

ОПТИЧЕСКИЕ ВЫЧИСЛЕНИЯ

Под ред. Р. Арратуна

М.: Мир, 1993.—441 с.

Сборник статей специалистов из США, ФРГ и Японии посвящен различным аспектам цифровых вычислений, осуществляемых оптическими и оптоэлектронными методами. Рассмотрены вопросы реализации параллельных вычислений оптическими методами, принципы построения вычислительных систем на основе волоконно-оптических линий, применения оптических методов для проведения символьных вычислений, в том числе для задач искусственного интеллекта. Сделана попытка на высоком научном уровне ознакомить читателей сразу со всем комплексом проблем, связанных с разработкой оптических цифровых вычислительных систем.

Для научных работников, занимающихся оптической обработкой информации, разработкой оптических процессоров и оптических вычислительных систем. Может быть использован как учебное пособие студентами и аспирантами, специализирующимися в области фотоники, оптики, радиофизики и информатики.

СОДЕРЖАНИЕ

Предисловие редактора перевода	5
Предисловие редактора серии	7
Предисловие	8
Об авторах	12
ЧАСТЬ I. ПРОСТРАНСТВЕННЫЕ МОДУЛЯТОРЫ СВЕТА	14
Глава 1. Магнитооптические модуляторы света. Б. Хилл	14
1.1. Введение	14
1.2. Принцип действия магнитооптических модуляторов	16
1.3. Магнитные свойства ферритмагнитных гранатов и механизмы переключения	18
1.4. Оптические свойства магнитооптических гранатовых пленок	26
1.5. Конструкция магнитооптических модуляторов	35
1.6. Обсуждение результатов и перспектив	48
Глава 2. Оптические бистабильные устройства и логические элементы. Н. Пейгамбарян, Дж. Джуэлл	52
2.1. Введение	52
2.2. Принципы работы бистабильных устройств	53
2.3. Требования к характеристикам нелинейных устройств и материалам	56
2.4. Причины возникновения нелинейных оптических эффектов в полупроводниках	57
2.5. Современное состояние работ по оптическим бистабильным устройствам	58
2.6. Оптические логические элементы Фабри — Перо (ОЛЭФП)	69
2.7. Устройства на собственном электрооптическом эффекте (СЭОУ)	67
2.8. Факторы, ограничивающие предельные величины коэффициентов усиления, разветвления по выходу, объединения по входу	69

2.9. Концепции архитектуры	71
2.10. Предельные параметры и вычислительные возможности ОБУ	72
2.11. Три-стабильность и многозначная логика	74
2.12. Выводы	74
Глава 3. Компоненты оптического процессора на основе приборов с зарядовой связью (ПЗС). Б. Е. Бёрк, Дж. Хиггинс	76
3.1. Введение	76
3.2. Примеры электрооптической обработки сигналов (ЭООС)	78
3.3. Принципы работы ПЗС	83
3.4. Применение ПЗС в качестве выходных устройств ЭООС	91
3.5. Адресуемые ПЗС-пространственные модуляторы света (ПЗС-ПМС)	97
3.6. Выводы	111
ЧАСТЬ II. МНОГОЗНАЧНАЯ И ПОРОГОВАЯ ЛОГИКА	113
Глава 4. Многозначная логика в оптических вычислениях. М. Коннер, Д. Эйхманн	113
4.1. Введение	113
4.2. Проблема полноты в двоичной логике	114
4.3. Многозначные логические схемы, основанные на системе счисления в остаточных классах (ССОК)	117
4.4. Оптические многозначные логические ССОК-процессоры	123
4.5. Проблема полноты в многозначной логике	132
4.6. Краткое изложение и выводы	139
Глава 5. Пороговое кодирование и взвешивание в оптических вычислениях. С. С. Густафсон	140
5.1. Введение	140
5.2. Устройства с внешним пороговым кодированием	146
5.3. Устройства с внутренним пороговым кодированием	154
5.4. Заключение	160
Глава 6. Многозначная пороговая логика. К. Морага	162
6.1. Введение	162
6.2. Многозначная пороговая логика	163
6.3. Мультилинейные разделяющиеся функции	172
6.4. Перспективные разработки	177
ЧАСТЬ III. СИСТОЛИЧЕСКИЕ ПРОЦЕССОРЫ И ОПТИЧЕСКИЕ ЛОГИЧЕСКИЕ МАТРИЦЫ	182
Глава 7. Оптическое умножение матриц. С. Кэртрайт	182
7.1. Введение	182
7.2. Цифровая оптика	183
7.3. Рабочие характеристики	204
7.4. Заключение	214
Глава 8. Архитектуры клеточной логики. Т. Ятаган	216
8.1. Введение	216
8.2. Клеточная логика	218
8.3. Оптическая двоичная логика и архитектура МКМД	225

8.4. Оптическая реализация	229
8.5. Заключение	236
Глава 9. Волоконно-оптические программируемые логические матрицы. Р. Арратун	237
9.1. Введение	237
9.2. Характеристики волоконно-оптических ПЛМ	245
9.3. Проблемы создания логических схем для волоконно-оптических систем	255
9.4. Перестраиваемые волоконные системы	263
9.5. Выводы	267
ЧАСТЬ IV. СИМВОЛЬНЫЕ ВЫЧИСЛЕНИЯ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ	271
Глава 10. Оптика и символьные вычисления. Дж. А. Нефф, Б. Дж. Кашнер	271
10.1. Введение	271
10.2. Что такое искусственный интеллект?	274
10.3. Функциональные возможности символьных вычислений	293
10.4. Архитектуры символьных вычислений	332
Глоссарий	357
Глава 11. Оптические межэлементные соединения для символьной и цифровой обработки в реальном времени. А. Макаулай	364
11.1. Введение	364
11.2. Роль параллелизма и символьной обработки в вычислениях проводимых в реальном времени	365
11.3. Преимущества перекрестных оптических схем	369
11.4. Оптический процессор с перекрестной схемой	374
11.5. Получение логических выводов с помощью оптического процессора с перекрестной схемой	382
11.6. Цифровая обработка сигналов с помощью оптического мультипроцессора с перекрестной схемой	386
11.7. Случай одновременного выполнения символьных и цифровых вычислений, пример речевой системы	391
11.8. Выводы	404
Литература	407

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Адресация перекрестная 39	— — — — — и многими потоками
Архитектура вычислений временной свертки с временным интегрированием 189	данных (МКМД) 224
— — — — — пространственным интегрированием 189	— с одним потоком команд и многими потоками данных (ОКМД) 224
— со многими потоками команд и одним потоком данных (МКОД) 224	— — — — — и одним потоком данных (ОКОД) 224
	— фон <i>Неймана</i> 218
	— чисто оптическая 346

Архитектуры вычисления свертки
оптическими методами 186
— гибридные второго поколения 269
База знаний 321
Бистабильные резонаторы *Фабри* —
Перо 53
Время переноса заряда 93
Выбеливание 105
Вывод логический 289
Вычисление разности в кадре 80
Вычисления символные 271
Вычислительные возможности 72
Гиперповерхность 177
Граф направленный 284
Группа *Абелева* p -элементарная 134
Декодер 240
Декодер высших порядков 257
Домены цилиндрические магнитные
Запоминающее устройство
динамическое с произвольной
выборкой (ДЗУПВ) 253
— — статическое с произвольной
выборкой (СЗУПВ) 253
Знания 275
Зрение техническое 294
Изменение временного масштаба
динамическое 298
Информация неточная 290
Каскад двухканальный 221
— *Мэйтры* 222
— с большим числом входов 221
Компьютеры клеточные 219
Корреляция 364
Коэффициент линейного
прогнозирования исходов 364
— объединения по входу волоконно-
оптических устройств 245
— — — — оптических бистабильных
устройств 69
— разветвления по выходу
волоконно-оптических
устройств 245
— — — — электронных логических
вентилей 340

Логика клеточная 218
— — голографическая 231
— пороговая p -значная 164
— — троичная 163
— программируемая 291
Матрица *Вандермонда* 118
— клеточная *Минника* 222
— логическая программируемая 240
— — — динамически 266
— — — оптоэлектронная 240
— мажоритарности 221
— микрокомпьютеров 334
— микролинз 235
— систолическая 78
Метод максимальной энтропии 389
— теневой энтропии 226
— логической минимизации 256
— — — ESPRESSO 256
— — — *Квина* — *Маккласки* 256
— — — MINI 256
— — — *Шейнманна* 256
Множество полное 115
— слабое полное 116
Моделирование авторегрессионное
364
Морфология 300
Мультимножество 173
Обработка данных на уровне вентиля
183
— — — — процессора 183
— — — — регистра 183
— конвейерная 350
— параллельная 333
— списков 272
Ограничение на плотность упаковки
246
— — рассеиваемую мощность 246
Операция корреляции 273
— поиска 273
— сравнения 273
Оптические логические элементы
Фабри — *Перо* 59
Отношение бинарное 133
— *Псалтиса* 210

— центральное 135
Отношения порядка 133
— эквивалентности нетривиальные 135
— 4-арные 127
Память ассоциативная 240
— прямого доступа 240
Параллелизм оптических вычислений 141
Переключатель перекрестный
волоконно-оптический 263
Переключение терромагнитное 36
Перекрестная сеть межэлементных соединений 240
Перенос заряда 86
ПЗС структура 76
— — на барьерах *Шоттки* 83
— — на GaAs 95
— — со скрытым каналом 85
— — с резистивным затворным слоем 90
Поиск перебором в глубину 284
— — — ширину 284
— от фактов к цели 284
— — цели к фактам 284
Понимание речи 297
— — непрерывной 298
— естественного языка 313
Поток данных 224
Прагматика 301
Предикат 281
Преобразователь аналого-цифровой 78
Произведение внешнее 201
— мощности на ширину полосы пропускания 254
— числа межэлементных соединений на ширину полосы пропускания 248
Производительность вычислений 241
Пространственный модулятор света на эффекте электропоглощения в ПЗС 98

Процессор вычисления свертки
систолический
акустооптический двоичный 195
— матричный систолический
оптический 193
— специальных функций 353
Разделение обобщенное 177
Распознавание отдельных слов 298
— речи символично-цифровое 393
Рассуждения 276
Свертка в фурье-плоскости 187
Семантика 295
Сети соединений межэлементных 334
Символьные подстановки 71
Синтаксис 295
Система счисления в остаточных классах 117
— экспертная 295
— — основанная на правилах 363
— нейронная 332
— продукции образец — действие 279
Сложность вычислений 241
— функциональная 261
Соединения межэлементные
оптические плата — плата 237
— — — чип — чип 237
Спектроанализатор
акустооптический 80
Структуры данных динамические 290
— — символьные 271
— квантоворазмерные 57
Таблицы перекодировочные
голографические 127
— — в клеточной логике 233
Теорема *Ферма* 122
Ток темновой 89
Триггер J — K 159
Три-стабильность 74
Умножитель двухразрядный 147
— матрицы на вектор 193
— — — матрицу 195
— -сумматор 155

Ускоритель 352
Устройство бистабильное оптическое
69
— вычислительное
быстродействующее
биполярное некогерентное на
светоделительном кубе
(RUBIC) 198
— запоминающее постоянное (ПЗУ)
240
— — с произвольной выборкой
(ЗУПВ) 240
— логическое программируемое
пользователем в зависимости от
условий эксплуатации
(ЛУПУЭ) 253
— на основе деформируемых зеркал
378
Фактор заполнения 91
Ферромагнетизм 19

Фильтрация, согласованная в
клеточной логике 235
Фонетика 300
Функция логическая элементарная
линейная 115
— — — монотонная 115
— — — самодвойственная 115
— — — сохраняющая 0 и 1 115
— разделяющаяся линейная 173
— мультилинейная 173
Частота появления ошибок 245
Число логических выводов в секунду
(липе) 272
Эвристика 283
Экситоны 57
Электромиграция 46
Элементы знания 271
Эффект *Фарадея* 26
— *Франца—Келдыша* 98
— электропоглощения в квантовых
ямах 105

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Появление этой книги симптоматично, поскольку интерес к оптической обработке информации вызвал всплеск надежд и работ сразу же после появления лазеров с высокой когерентностью. Однако реализация оптических процессоров столкнулась с целым рядом практических трудностей. К ним относятся создание пространственно-временных модуляторов света (т. е. способов ввода информации в оптическую систему), разработка новых типов сред для регистрации информации, реализация архивной и динамической систем памяти и др. Возникающие на пути практической реализации оптических процессоров проблемы в какой-то мере охладили восторг, появившийся после осознания достоинств оптической обработки информации, но не остановили эти исследования.

В настоящее время уже создано достаточно большое число различных видов оптических процессоров, назначение, разумность и возможность применения которых ни у кого не вызывает сомнений. Примером достаточно хорошо проработанных в фундаментальном плане и с точки зрения реализации устройств являются магнитооптические процессоры, описанию которых посвящена первая глава книги. Хотелось бы заметить, что это далеко не единственный вид практически реализованных оптических процессоров. Как за рубежом, так и в нашей стране имеется большое число энтузиастов, усиленно ведущих работу в этом направлении.

Данный сборник написан коллективом специалистов различного профиля, и отчетливо видно, что каждый из них — специалист высокого класса в своей области. Это, несомненно, большое преимущество данного изделия. Другая его особенность состоит в том, что возможности, достоинства и пути реализации преимуществ оптической обработки информации (а это прежде всего параллельность обработки) теперь уже волнуют и вызывают глубокие размышления не только у физиков, но и у специалистов в области вычислительной техники. Это весьма знаменательно! Однако большое число авторов, яв-

ляющихся специалистами в разных областях, неизбежно использовали разный стиль изложения и различную терминологию и трактовку одних и тех же словосочетаний. Это, например, касается таких понятий, как «аналоговые вычисления» и «символьные вычисления», появляющиеся соответственно в начале книги и при обсуждении проблем искусственного интеллекта (в части IV). Редакторы и переводчик старались сохранить индивидуальность каждого автора.

Советским читателям, с нашей точки зрения, будут особенно интересны главы, посвященные многозначной логике и путям ее реализации в оптике, систолическим процессорам и оптическим логическим матрицам, оптической обработке и искусственному интеллекту. Интересен своим практическим подходом раздел, в котором рассматриваются волоконно-оптические программируемые логические матрицы. Остается только сожалеть, что в книгу не включены материалы о процессорах нейронного типа. Для лиц, занимающихся созданием вычислительных машин, очень полезны разделы, посвященные оптическим межэлементным соединениям для цифровой и символьной обработки в реальном времени.

В целом книга полезна для широкого круга читателей: студентов, аспирантов, инженеров, научных работников как в области оптики, голографии, оптической обработки информации, так и в области проектирования электронных вычислительных машин. Поскольку книга написана в духе обсуждения путей и возможностей, она не только дает информацию, но в еще большей мере пищу для размышлений и поиска как способов реализации параллельности обработки в гибридных оптико-электронных схемах и чисто оптических вычислительных машинах, так и методов создания специализированных оптических процессоров.

Ю. А. Быковский

За последнее десятилетие оптическая наука, техника и технология быстро развились, так что в настоящее время оптическая техника стала самостоятельной и важной дисциплиной. Данная серия посвящена обсуждению проблем оптической техники на том уровне, который будет полезен тем, кто работает в данной отрасли или делает попытки конструировать системы либо основанные на оптических устройствах, либо включающие в себя значительные оптические подсистемы. Основная ее цель состоит не в том, чтобы представить детальные монографии по узким проблемам, но в рассмотрении материала на том уровне, который сразу же принесет пользу для учено-практика и инженера. Эти издания не являются научно-исследовательскими монографиями, хотя мы ожидаем, что читатели, занимающиеся научными исследованиями в области оптики, найдут их весьма ценными.

Издания данной серии охватывают разные проблемы быстро развившейся области оптической техники. Разработки, приведшие к бурному развитию этого направления, включают вопросы, посвященные лазерам и их многочисленным техническим и промышленным применениям, новым оптическим материалам, градиентной оптике, электро- и акустооптике, волоконной оптике и связи, оптическим вычислениям и распознаванию образов, считыванию, записи и хранению оптической информации, биомедицинской измерительной технике, промышленным роботам, интегральной оптике, системам инфракрасного и ультрафиолетового диапазонов и т. д. Поскольку оптическая промышленность в настоящее время является одной из основных развивающихся отраслей, то этот список, несомненно, станет еще более обширным.

Брайан Дж. Томпсон
Рочестерский университет
Рочестер, шт. Нью-Йорк

Во второй половине двадцатого века в технических науках произошли резкие изменения, в результате чего наступила эпоха информатики. Возможно, основной движущей силой этих изменений явилось появление электронных цифровых вычислительных машин. Технические возможности ЭВМ хотя и впечатляют, тем не менее являются ограниченными. Развитие программных средств и архитектуры, вероятно, еще некоторое время будет продолжаться, однако уже сейчас ощущаются практические ограничения возможностей субмикронной технологии и электронных систем межэлементных соединений¹⁾, предназначенных для цифровых логических элементов и являющихся основными составными частями любого цифрового компьютера. Наряду с этим существует оптическая методика, позволяющая обойти некоторые из этих ограничений. Эта область называется *оптическими вычислениями*.

Область оптических вычислений быстро расширяется. В историческом плане на эту область оказали сильное влияние попытки выгодно использовать изначальную способность оптических систем выполнять аналоговые умножение и сложение. Несмотря на фанфарный звон, возникший вокруг аналоговых оптических вычислений, цифровые оптические вычисления без лишнего шума применялись во многих существующих вычислительных системах для обеспечения соединений уровней система — система, плата — плата, чип — чип. В самом деле, уклон в сторону цифровых оптических вычислений определялся попытками воспользоваться преимуществами при использовании особых свойств световых пучков в соединениях. Так как возможности соединений тесно связаны со способностью системы выполнять параллельные вычисления, то оптика предоставляет естественный механизм, способствующий разработке параллельной архитектуры компьютеров. Хотя соединения имеют

¹⁾ Далее — *соединения*. В литературе также используется термин *межсоединения*. — *Прим. ред.*

большое значение на всех уровнях архитектуры компьютера, необходимо также рассмотреть вопрос разработки логических элементов. По этой причине рассматриваются и другие аспекты исследований цифрового оптического компьютера, связанные с реализацией логических операторов, более мощных по сравнению с традиционными элементарными булевыми операторами. Устройства этого вида включают как пороговые логические, так и многозначные логические вентили. Другой весьма быстроразвивающейся разработкой в области элементной базы является появление сверхбыстродействующих оптических бистабильных элементов, на которых теоретически можно выполнить множество логических операций и действий с памятью.

В настоящей книге особое внимание уделено цифровым аспектам оптических вычислений, так же как и новой увлекательной области, называемой *символьными вычислениями*. Существует мнение, что грядущее шестое поколение компьютеров будет состоять из символьных, а не цифровых машин. Этот тип компьютеров сможет давать приближенные ответы на задачи, которые не требуют точных решений. На уровне чипов эти машины будут характеризоваться способностью выполнять поиск, сортировку данных и реализовать ассоциативные связи (в базах данных) с исключительно высокими скоростями. На системном уровне возможности этих компьютеров будут расширены до способности получать логические выводы. Такие компьютеры будут способны работать с эвристическими аргументами, смогут включать в себя семантическую память, позволяющую им говорить, например, по-английски. Следует заметить, что символьные вычисления отличаются от подстановок в символьном виде в том, что последние представляют собой алгоритмический подход, ориентированный на набор точно определенных правил или логических операций, применяемых в качестве параллельных «окон» в большом блоке данных. Этот подход в основном эквивалентен режиму работы с одним потоком команд и многими потоками данных (ОКМД), за исключением того, что правила (команды) могут быть скорректированы в интервалах между проходами, чтобы обеспечить выполнение последовательных алгоритмов. В противоположность этому символьные вычисления включают в себя операции более высокого уровня и ориентированы на эвристические подходы.

В данную книгу не включена обзорная глава, специально посвященная нейронным сетям, отчасти потому, что эта область очень быстро развивается, и отчасти по той причине, что большинство таких систем, реализованных в оптике, являются либо аналоговыми, либо псевдоаналоговыми. В первой части книги рассмотрены пространственные модуляторы света. В циф-

ровом варианте эти устройства по существу являются матрицами оптических переключателей, которые либо пропускают, либо блокируют свет. Большие матрицы таких переключателей дают возможность создания перестраиваемых оптических соединений, весьма универсального элемента в усовершенствованных системах параллельной обработки и распределенных структурах. В этой книге не предпринималось попыток дать исчерпывающий обзор всех видов пространственных модуляторов света, поскольку это потребовало бы написания нескольких томов. Рассмотрены несколько типичных примеров перспективных устройств данного типа. Б. Хиллом (Bernard Hill) сделан обзор магнитооптических устройств, в то время как Н. Пейгамбаряном (Nasser Peyghambarian) и Дж. Джуэллом (Jack Jewell) рассмотрены оптические бистабильные устройства. Как уже ранее обсуждалось в литературе, бистабильные устройства, несомненно, представляют собой нечто большее, чем просто переключатели, и теоретически способны выполнять разнообразные логические операции и операции с памятью. В гл. 3 Дж. Хиггинс (John Higgins) и Б. Бёрк (Barry Burke) дают исчерпывающий обзор компонент оптических процессоров на основе приборов с зарядовой связью (ПЗС).

В части II особое внимание уделено многозначной и пороговой логике. Элементарные логические операции, связанные с этими устройствами, дают возможность заменить большое число простых вентилей набором более мощных логических элементов. Этот подход имеет дополнительное преимущество в том, что позволяет снизить число межэлементных соединений, необходимых для конкретного уровня сложности вычислений. В главе, написанной М. Коннером (Michael Conner) и Дж. Эйхманном (George Eichmann), обсуждается многозначная логика, в то время как С. Густафсон (Steven Gustafson) рассматривает пороговую логику. К. Морага (Claudio Moraga) в своем математически строгом изложении объединяет оба этих подхода к проблеме многозначной пороговой логики.

В части III рассматриваются систолические процессоры и матрицы оптических логических элементов. В главе, посвященной систолическим процессорам, С. Кэртрайт (Steven Cartwright) делает обзор ряда работ по оптике, в которых предпринимаются попытки осуществить параллельные цифровые вычисления с помощью аналоговых методик. В гл. 8 Т. Ятагаи (Toyohiko Jatagai) обсуждает чисто цифровые методы оптических вычислений, основанных на клеточной логике. Данный подход теоретически допускает работу как в ОКМД, так и в режиме с многими потоками команд и многими потоками данных (МКМД). В написанной мною главе рассмотрены фундаментальные физические ограничения при реализации оптической логики на основе волоконно-оптических программируемых

логических матриц. В настоящей главе с позиций чисто комбинаторной логической системы также обсуждается различие между производительностью и сложностью вычислений.

В IV части данной книги содержатся две главы, посвященные символьным вычислениям. В гл. 10 Дж. Нефф (John Neff) и Б. Кашнер (Brian Kushner) представили превосходный учебный материал по проблемам искусственного интеллекта и символьных вычислений. В последней главе А. Мак-Аулэй (Alastair McAulay) подчеркивает преимущества перестраиваемых оптических систем и дает глубокое описание того, как основанные на правилах логические выводы можно получить с помощью оптических перекрестных процессоров.

Главы этой книги содержат оригинальные и методические сведения, позволяющие использовать ее на семинарах для аспирантов, и в практических научных исследованиях. Я надеюсь, что данный материал вызовет у читателей интерес как к цифровым, так и к символьным оптическим вычислениям.

Раймонд Арратун

- Раймонд Арратун** (Raymond Arrathoon). Отделение электро- и вычислительной техники, Государственный университет им. Уэйна, Детройт, шт. Мичиган.
- Барри Е. Бёрк** (Barry E. Burke). Лаборатория им. Линкольна, Массачусетский технологический институт, Лексингтон, шт. Массачусетс.
- Стивен Кэртрайт** (Steven Cartwright)¹⁾. Исследовательский институт Университета г. Дайтона, Отделение прикладной физики, Университет г. Дайтона, Дайтон, шт. Огайо.
- Мишель Коннер** (Michael Conner). Отделение электротехники, Городской колледж Университета г. Нью-Йорка, Нью-Йорк, шт. Нью-Йорк.
- Джордж Эйхманн** (George Eichmann). Отделение электротехники, Городской колледж Университета г. Нью-Йорка, Нью-Йорк, шт. Нью-Йорк.
- Стивен С. Густафсон** (Steven C. Gustafson). Исследовательский институт, Отделение прикладной физики, Университет г. Дайтона, Дайтон, шт. Огайо.
- Джон Р. Хиггинс** (John A. Higgins). Центр исследований и разработки микрoeлектронных структур, Роквелл Интернэйшенэл, Саузенд Оукс, шт. Калифорния.
- Бернард Хилл** (Bernard Hill). Институт технической электроники, Технический университет г. Аахена, Аахен, ФРГ.
- Джек Л. Джуэлл** (Jack L. Jewell). АТТ Белл Лэборэториз, Холмдел, шт. Нью-Джерси.
- Брайан Дж. Кашнер** (Brian G. Kushner). Отделение информационных технологий, БДМкорпорэйши, Арлингтон, шт. Вирджиния.
- Аластер Д. Мак-Аулай** (Alastair D. McAulay). Отделение вычислительной науки и техники, Государственный университет им. Райта, Дайтон, шт. Огайо.

¹⁾ В настоящее время — Исследовательский институт окружающей среды оз. Мичиган, Техническое бюро г. Дайтона, Дайтон, шт. Огайо.

- Клаудио Морага** (Claudio Moraga). Отделение вычислительной техники, Университет г. Дортмунда, Дортмунд, ФРГ.
- Джон А. Нефф** (John A. Neff). Отдел электроники, Агентство оборонных исследовательских проектов, Арлингтон, шт. Вирджиния.
- Насер Пейгамбарян** (Nasser Peyghambarian). Центр оптических исследований, Университет шт. Аризона, Таксон, шт. Аризона.
- Тойохико Ятагаи** (Toyohiko Jatagai). Институт прикладной физики, Университет г. Цукубэ, Цукубэ, Ибараки, Япония.

ЧАСТЬ I

ПРОСТРАНСТВЕННЫЕ МОДУЛЯТОРЫ СВЕТА

Глава 1

МАГНИТООПТИЧЕСКИЕ МОДУЛЯТОРЫ СВЕТА

*Б. Хилл, Институт технической электроники,
Технический университет г. Аахена, Аахен, ФРГ*

1.1. Введение

За последние годы магнитооптические модуляторы на основе кристаллов граната стали представлять особый интерес с точки зрения создания оптоэлектронных устройств. В области оптической обработки сигналов просматриваются многочисленные приложения, связанные с сопряжением электрических и оптических сигналов, созданием пространственных фильтров, перестраиваемых в фурье-плоскости, с реализацией логических функций. Другие применения обнаруживаются в оптических системах, используемых в литографии и системах оптической связи.

Детальное изучение магнитооптических пленок граната началось в середине 1960-х годов [1, 2]. В это время на кристаллы граната возлагали большие надежды, как на основной материал для систем оптической памяти [3, 4]. Работа, проводившаяся в данном направлении, была связана с развитием систем памяти на цилиндрических магнитных доменах (ЦМД), в которых использовалось управляемое перемещение магнитных доменов. Чтобы сделать пути продвижения доменов видимыми для наблюдения в микроскопе, использовался магнитооптический эффект Фарадея при пропускании света. Этот эффект, однако, в применявшихся тогда материалах был очень слабым. В 1972 г. было обнаружено, что введение висмута в кристаллы граната сильно увеличивает эффект Фарадея в оптической области спектра. Это открытие наряду с известными методиками работы с ЦМД послужило толчком к появлению первой разработки в направлении создания магнитооптического модулятора — дисплея на ЦМД [6, 7]. Квадратная пластина граната на основе железа и гадолиния, замещенного висмутом, размещалась в оптической схеме с поляризованным светом. В пластине возникала определенная структура цилиндрических магнитных доменов. Домены, имеющие противоположные направления намагниченности, при пропускании света наблюдались как черные и белые точки. Передача полезного сигнала прошедшим световым потоком была, однако, очень мала

вследствие весьма малых размеров ЦМД по сравнению с их окружением, что было характерно для структур ЦМД. Эта работа соответственно не была завершена.

На основе разработок в области материалов для оптической памяти [3, 4] несколькими годами позже был предложен другой подход, заключающийся в изготовлении островковой структуры гранатовой пленки на специальной подложке [8, 9]. В этой структуре каждый островок содержит локально фиксированный магнитный домен, намагниченность которого переключается термомагнитным способом, т. е. подводом тепла и импульса магнитного поля. Так как островки отделены друг от друга только маленькими канавками, то большая часть площади элемента может быть использована в оптической схеме и соответственно коэффициент передачи оптического полезного сигнала может быть значительно улучшен.

Аналогичной разработкой, основанной на иной схеме переключения, является структура, названная создателями *лайт-мод*¹⁾ [10, 11]. В этом устройстве, чтобы избавиться от подачи тепловых импульсов, зарождение доменов с противоположным направлением намагниченности инициируется в ионно-имплантированных центрах. Переключение здесь достигается только приложением управляющего магнитного поля. Такой подход позволяет добиться больших скоростей переключения, чем при термомагнитном переключении.

В следующих разделах кратко обрисованы вопросы подбора материалов и основные характеристики гранатов как основного материала для магнитооптических модуляторов. Затем обсуждаются различные конструкции и методы адресации магнитооптических модуляторов наряду с кратким изложением экспериментальных результатов и предельных возможностей. В заключение будут даны некоторые оценки будущих разработок с точки зрения их применения для оптической обработки сигналов.

1.2. Принцип действия магнитооптических модуляторов

Магнитооптический модулятор состоит из подложки из немагнитного монокристаллического граната, на которую нанесена тонкая монокристаллическая пленка ферритмагнитного граната (рис. 1.1). Эта пленка обладает спонтанной намагниченностью, всегда направленной перпендикулярно к плоскости самой пленки. Кроме того, она также обладает оптическим эффектом Фарадея. По причинам, указанным в разд. 1.3, эта пленка

¹⁾ Русскоязычный эквивалент аббревиатуры *lightmod*, предложенной в [10] в качестве названия устройства, представляется менее благозвучным по сравнению с английским вариантом. — *Прим. перев.*

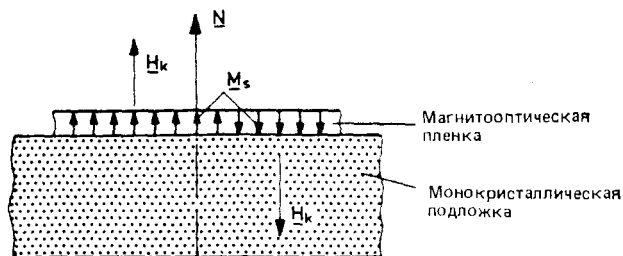


Рис. 1.1. Пленка феррита-граната, нанесенная на монокристаллическую немагнитную подложку. Ферромагнитная гранатовая пленка, обычно имеющая толщину 5—10 мкм, эпитаксиально выращивается на монокристаллической подложке (толщиной ~ 500 мкм) и обладает спонтанной намагниченностью насыщения M_s , которая обусловлена одноосевой анизотропией, ориентированной либо параллельно, либо антипараллельно по направлению нормали N к пленке. Анизотропия теоретически описывается полем анизотропии H_k .

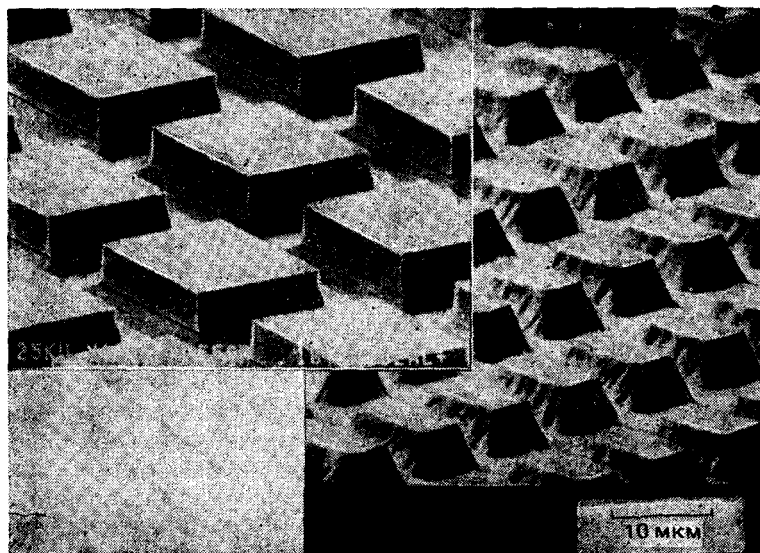


Рис. 1.2. Структура магнитооптической пленки. На магнитооптической пленке, нанесенной на подложку фотолитографическими методами, вытравлены отдельные переключательные ячейки. Каждая ячейка имеет площадь от 10 до 100 мкм². На рисунке показаны как 50-микронные структуры (верхний рис.), так и 10-микронные структуры (нижний рис.) [11].

разделена на отдельные островки, каждый из которых является основой для переключающей ячейки (рис. 1.2).

Как показано на рис. 1.3, в зависимости от направления собственной намагниченности плоскость поляризации линейно-поляризованного света, прошедшего через переключающую ячейку, вращается либо по часовой, либо против часовой стрелки. При изменении направления собственной намагниченности на противоположное направление фарадеевского вращения изменяется с направления вращения по часовой стрелке на направление вращения против часовой стрелки и наоборот. Расположенный за гранатовым элементом поляризатор-анализатор не пропускает свет для одной плоскости поляризации и соответственно пропускает его для другой плоскости поляризации. Таким образом, имеется контраст между потоками света, проходящими через ячейки с противоположными направлениями собственной намагниченности.

Существуют различные методы переключения намагниченности ячейки. Для того чтобы понять механизм переключения и получить представление о существенных параметрах пере-

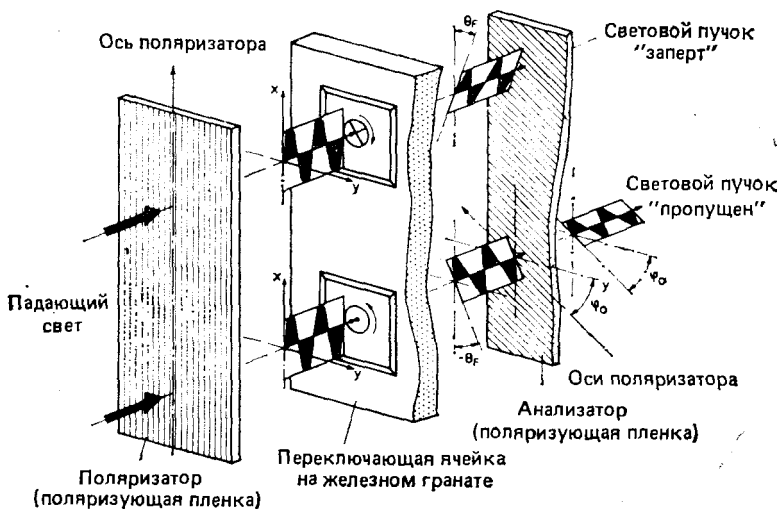


Рис. 1.3. Оптическая схема магнитооптического модулятора. Гранатовый элемент размещен между поляризатором и поляризатором-анализатором (экспериментально реализуемых в виде поляризирующих пленок). Из поляризатора выходит линейно-поляризованный свет с поляризацией в вертикальной плоскости. Для света, прошедшего через ячейку феррита-граната, плоскость поляризации повернута по часовой либо против часовой стрелки. Анализатор настроен таким образом, что он не пропускает свет, прошедший через верхнюю гранатовую ячейку, намагниченность которой параллельна направлению пучка света. Но свет, прошедший через нижнюю ячейку, имеющую противоположное направление намагниченности, анализатор пропускает.

ключения и предельных возможностях, следует рассмотреть некоторые фундаментальные свойства пленок ферритов-гранатов. Соответственно в следующем разделе основное внимание сконцентрировано на свойствах пленок ферритов-гранатов, подробном описании конструкций, способах адресации переключающей ячейки и используемых переключающих механизмах.

Необходимо также более детально рассмотреть оптические свойства, чтобы понять, чем ограничиваются контраст при оптическом переключении и коэффициент передачи полезного сигнала устройства. Указанные проблемы подробно обсуждаются в разд. 1.3.

1.3. Магнитные свойства ферритмагнитных гранатов и механизмы переключения

В общих чертах гранаты представляют собой смесь кислорода с различными элементами, такими, как редкоземельные элементы и металлы. Ионы этих элементов образуют особый вид кристаллических решеток, называемый решеткой типа граната. Хорошо известные природные гранаты — пироп и альмандин — состоят из кальция, алюминия, кремния и небольшого количества железа, марганца или, например, титана. Обширный обзор по гранатам представлен в [12].

Существенно, что часть названных компонентов может быть заменена другими химическими элементами, что значительно расширяет число возможных комбинаций. Обсуждаемые в данной главе искусственные магнитооптические гранаты состоят из переходных металлов — железа и гадолиния, при этом часть этих химических элементов замещена висмутом и галлием. Типичный состав магнитооптического граната следующий: $Gd_{3-x}Bi_{5-y}Ga_yO_{12}$. Степени замещения (коэффициенты x и y) сильно влияют на магнитные и оптические свойства и используются для модификации оптических свойств материала, а, следовательно, оптимального подбора материалов.

Монокристаллические искусственные гранаты выращивают в бульках из расплава по методу Чохральского. В магнитооптическом модуляторе используются тонкие (5—10 мкм) пленки гранатов. Тонкие пленки гранатов выращиваются эпитаксиально на немагнитных гранатовых подложках при погружении их в жидкий расплавленный материал. Подложки вырезаются из булек с наращенным материалом и полируются. В настоящее время стандартными являются подложки размером до 7,5 см (толщиной ~500 нм). Чтобы получить высококачественную монокристаллическую пленку без внутренних напряжений, необходимо, чтобы периоды решеток подложки и пленки (обычно 1,246 нм) совпадали. Выполнение данного условия иногда

приводит к усложнению состава подложки, как, например, используемого в [13] $Gd_{3-v}Ca_vGa_{5-v-2z}Mg_zZr_{v+z}O_{12}$ в сочетании с вышеупомянутым гадолиниево-галлиевым гранатом.

1.3.1. Ферримагнетизм

В магнитооптических модуляторах используются гранатовые пленки ферримагнитного типа. Ферримагнетизм обусловлен магнитными дипольными моментами ионов железа и гадолиния, расположенных в определенных узлах решетки граната (например, ионы Fe^{3+} — в узлах октаэдра и тетраэдра, а Gd^{3+} — в узлах додекаэдра). Дипольные моменты, создаваемые спинами электронов в этих ионах, обусловлены квантово-механическими силами, ориентированными параллельно и антипараллельно друг к другу, в сумме дающими результирующую намагниченность (магнитный момент на единицу объема) M_s (рис. 1.1) [13].

В дополнение к этому имеется одноосная анизотропия, заставляющая результирующую намагниченность ориентироваться или параллельно, или антипараллельно нормали к пленке (рис. 1.1). Анизотропия возникает в основном в процессе роста вследствие деформации или напряжений в пленке. В теории магнетизма влияние анизотропии описывают полем анизотропии H_k . Предполагается, что это фиктивное магнитное поле присутствует в модели материала вместо физической анизотропии и определяется тем условием, что оно удерживает намагниченность, ориентированную вдоль выделенной оси с той же силой, как и анизотропия в реальных материалах. Оно ориентировано параллельно нормали N к пленке, если намагниченность имеет компоненту, параллельную нормали, и соответственно противоположное направление — в противном случае (см. рис. 1.1). В зависимости от состава материала поле анизотропии составляет по порядку величины от 10 до 1000 кА/м [14].

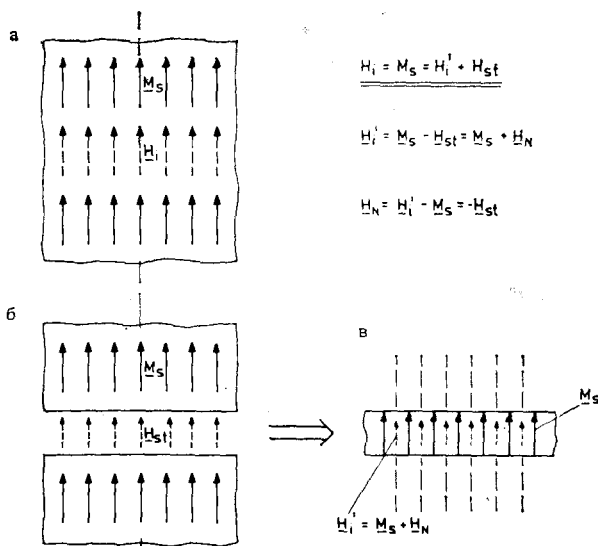
1.3.2. Формирование магнитных доменов

Параллельная или антипараллельная ориентация дипольных моментов в присутствии одноосной анизотропии вызывает появление замкнутых областей пленки, намагниченных одинаковым образом. Такая намагниченная область называется магнитным доменом.

Однако однородно намагниченная пленка, имеющая малую толщину по сравнению с ее площадью, создает сильное размагничивающее поле H_N (рис. 1.4). Если это поле превышает поле анизотропии H_k , то результирующее поле $H_N + H_k$ будет направлено противоположно намагниченности. В результате направление намагниченности изменяется на обратное. Таким

образом, пленка будет разделена на некоторое число доменов с противоположной намагниченностью. Для высококачественного материала в конечном итоге около половины площади пленки будет занято доменами с противоположным направлением намагниченности. Так как для очень тонких магнитооптических пленок, рассматриваемых в данной главе (см. рис. 1.4), величина \underline{H}_N составляет $|\underline{H}_N| \approx |\underline{M}_s|$, условием спонтанного разбиения в доменную структуру является неравенство $|\underline{H}_k| < |\underline{M}_s|$.

Образование структуры доменов снижает рассеянное поле вне пленки и локальное размагничивающее поле в пленке. Усредненное по площади A домена размагничивающее поле



$$\underline{H}_i = \underline{M}_s = \underline{H}_i' + \underline{H}_{st}$$

$$\underline{H}_i' = \underline{M}_s - \underline{H}_{st} = \underline{M}_s + \underline{H}_N$$

$$\underline{H}_N = \underline{H}_i' - \underline{M}_s = -\underline{H}_{st}$$

Рис. 1.4. Внутреннее поле и размагничивающее поле. Внутреннее поле в однородно намагниченном и неограниченном образце (а) с насыщенной намагниченностью \underline{M}_s определяется соотношением $\underline{H}_i = \underline{M}_s$. Если из образца вырезана тонкая пластина (б), намагниченность \underline{M}_s оставшейся части образца (б) создает в щели рассеянное поле \underline{H}_{st} . Намагниченность \underline{M}_s в вырезанной пластине порождает рассеянное поле вне пластины и уменьшенное внутреннее поле \underline{H}_i' . Поскольку линейная суперпозиция полей в случаях (б) и (в) должна давать такое же поле, как в случае (а), внутреннее поле \underline{H}_i' уменьшено по сравнению с $\underline{H}_i = \underline{M}_s$ на величину \underline{H}_{st} . Эта величина называется размагничивающим полем $\underline{H}_N = -\underline{H}_{st}$. Внутреннее поле \underline{H}_i' , таким образом, рассматривается как результат сложения намагниченности \underline{M}_s и размагничивающего поля \underline{H}_N . Если вырезанная пластина сделана очень тонкой, величина размагничивающего поля \underline{H}_N приближается к значению \underline{M}_s (в пределе $d \rightarrow 0$, $|\underline{H}_N|$ равняется $|\underline{M}_s|$).

$\langle H_N \rangle = (1/A) \int H_N dA$, так же как и средняя намагниченность $\langle M \rangle = (1/A) \int \overline{M_s} dA$, уменьшается почти до нуля (рис. 1.5). На рис. 1.6 показана типичная микроструктура пленки указанного типа, характеризующаяся структурой полосовых доменов.

Область между магнитными доменами, где намагниченность изменяет свое направление на обратное, называется доменной стенкой. Для образования доменных стенок требуется приложение энергии, так как при изменении направления намагниченности приходится преодолевать квантово-механические силы. Следовательно, размер доменной стенки не уменьшается до нуля, что приводило бы к полному исчезновению рассеянного поля и размагничивающего поля. Размер доменов определяется скорее минимумом соответствующих энергий магнитных полей: существует энергия стенки доменов и энергия

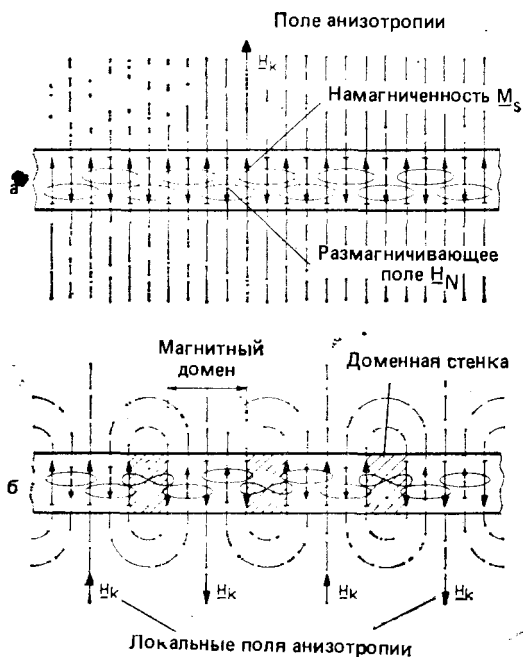


Рис. 1.5. Схема формирования доменов. В однородно намагниченной пленке (а) величина размагничивающего поля (см. рис. 1.4) приблизительно равняется намагниченности $|M_s|$: $|H_N| \approx |M_s|$. Если поле анизотропии $|H_k|$ меньше $|M_s|$, сильные размагничивающие поля заставляют намагниченность в половине материала пленки изменить направление на противоположное: формируется структура доменов с противоположным направлением намагниченности. В результате уменьшаются величины рассеянного поля извне пленки и локального размагничивающего поля в пленке. Пленка становится размагниченной.

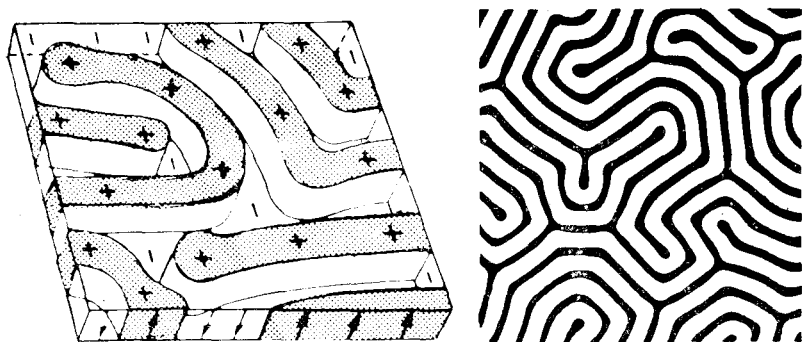


Рис. 1.6. Полосовые домены в магнитооптической пленке граната. Пленка, обладающая величиной поля анизотропии, меньшей, чем намагниченность насыщения \underline{M}_s , разделена на полосовые домены с типичной шириной в данном материале в 100 мкм. Домены становятся видимыми при освещении магнитооптической пленки поляризованным светом.

поля намагниченностей доменов. В предельном случае бесконечно малых доменов поля намагниченностей взаимно уничтожают друг друга, в то время как энергия стенок доменов достигает максимума (вся пленка занята «магнитными» стенками). Если, с другой стороны, пленка однородно намагничена, то энергия стенки доменов является нулевой, а энергия поля, определяемого \underline{M}_s , является максимальной. Равновесное состояние устанавливается между этими двумя предельными случаями.

1.3.3. Подвод внешнего магнитного поля и петля гистерезиса

1.3.3.1. Случай, когда намагниченность насыщения превышает поле анизотропии

Картина распределения магнитных доменов, спонтанно возникающих в магнитооптической пленке при $|\underline{H}_k| < |\underline{M}_s|$, является не слишком устойчивой. Малые изменения одной или более соответствующих магнитных сил или энергий, определяющих состояние равновесия, изменят распределение доменов.

Следовательно, картина распределения доменов также является чувствительной к приложенным внешним магнитным полям. Внешнее поле \underline{H} , приложенное по нормали к пленке, благоприятствует магнитным доменам, намагниченность которых параллельна этому полю, в то время как домены с противоположной намагниченностью сжимаются и уменьшают свои размеры.

В случае приложения сильного импульсного внешнего поля магнитные стенки между растущими и сжимающимися домена-

ми могут перемещаться весьма быстро. Экспериментально наблюдали скорости продвижения до 100 м/с (100 мкм за 1 мкс).

Начиная с размагниченного состояния магнитной пленки, зависимость $\langle M \rangle$ от H является почти линейной функцией (рис. 1.7). При $|\underline{H}| = H_s$ домены с противоположным направлением намагниченности коллапсируют и полностью исчезают. Пленка в данном состоянии является однородно намагниченной и насыщенной. H_s соответственно называют насыщающим полем. Увеличение приложенного поля более этого значения не вызовет дальнейшего изменения состояния материала.

Если, с другой стороны, уменьшать поле ниже значения H_s , насыщенное состояние остается стабильным до тех пор, пока H не станет меньше поля анизотропии, за вычетом размагничивающего поля (составляющего величину $\sim M_s$): $|\underline{H}| < < \underline{H}_k - \underline{M}_s |$. Тогда состояние намагниченности пленки опять

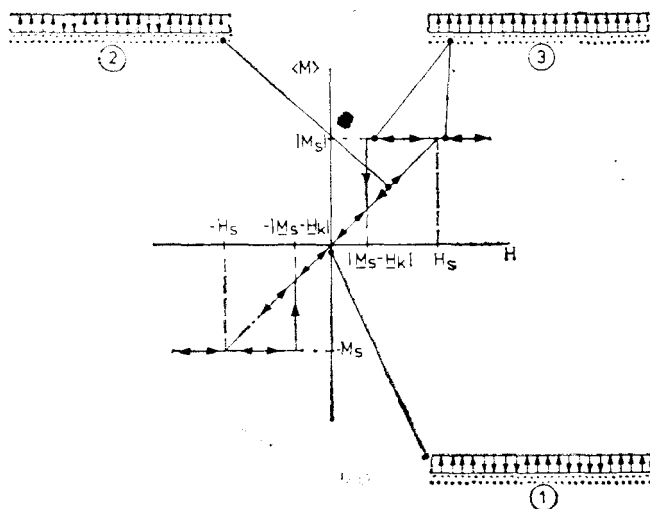


Рис. 1.7. Петля гистерезиса для случая $|\underline{H}_k| < |\underline{M}_s|$. В отсутствие внешнего поля размагничивающее поле \underline{H}_N размагничивает пленку (случай (1)). Для приложенного внешнего поля \underline{H} домены с направлением намагниченности, противоположным полю \underline{H} , сжимаются, а другие растут (случай (2)). При этом средняя намагниченность пленки $\langle M \rangle$ растет. При $|\underline{H}| \leq H_s$ пленка однородно намагничена и «насыщена», при этом намагниченность $\langle M \rangle = |\underline{M}_s|$. H_s называют полем насыщения (случай (3)).

Этот случай остается устойчивым, если \underline{H} уменьшено, и сумма приложенного поля и поля анизотропии превышает размагничивающее поле $|\underline{H}_N| = |\underline{H}_{N \text{ макс}}| = |\underline{M}_s|$. Если $\underline{H} + \underline{H}_k$ становится меньше M_s , пленка снова самопроизвольно дробится на домены. Критическая точка, в которой пленка становится неустойчивой, задается условием $|\underline{H}| \leq |\underline{H}_k - \underline{M}_s|$.

становится неустойчивым, и в пленке возникает доменная структура. Этот процесс требует затрат энергии на изменение направления намагниченности насыщения, на которую действует сила, создаваемая полем анизотропии, а также затрат энергии на формирование магнитной доменной стенки. Требуемая энергия может быть получена за счет высокой энергии рассеянного поля насыщенной пленки. Как только зарождаются домены с противоположным направлением намагниченности, в пленке опять самопроизвольно возникает доменная структура (рис. 1.7).

1.3.3.2. *Случай поля анизотропии, превышающего намагниченность насыщения*

Как указано выше, величина, равная разности поля анизотропии и намагниченности насыщения $|\underline{H}_k - \underline{M}_s|$, является существенной компонентой поля, определяющей устойчивость насыщенной и однородно намагниченной пленки. Начиная с изначально размагниченого состояния материала, приложение внешнего поля в любом случае приведет к сжатию доменов, имеющих намагниченность, противоположную этому полю, и $\langle M \rangle$ увеличивается пропорционально приложенному полю. При значении H_s материал является однородно намагниченным. Если затем приложенное поле уменьшают, то насыщенное состояние остается устойчивым настолько долго, насколько сумма приложенного поля и поля анизотропии $\underline{H}_k + \underline{H}$ превышает поле анизотропии $H_N = -M_s$. До этого момента материал ведет себя так, как указывалось в разд. 1.2.3.1. Однако, чтобы достичь неустойчивого состояния намагниченности с целью переключения, направление приложенного поля должно быть изменено на обратное и соответственно должно удовлетворять условию $H < -|\underline{H}_k - \underline{M}_s|$ (или $|H| > |\underline{H}_k - \underline{M}_s|$).

Теперь следует различать два случая [10]. В первом случае поле анизотропии за вычетом намагниченности насыщения считается меньшим, чем H_s : $|\underline{H}_k - \underline{M}_s| < H_s$. В данном случае пленка будет дробиться на домены, если величина обращенного по направлению поля \underline{H} превышает $|\underline{H}_k - \underline{M}_s|$, но является все еще меньшей, чем H_s . Это показано более детально на рис. 1.8. Зависимость средней намагниченности от приложенного поля образует типичную гистерезисную петлю, однако размагниченное состояние может быть вновь достигнуто, если следовать пунктирной линии на рис. 1.8.

С другой стороны, если разность поля анизотропии и намагниченности насыщения превышает поле насыщения, размагниченное состояние никогда не может быть опять достигнуто после того, как материал однажды был введен в насыщенное состояние (рис. 1.9). Этот случай характеризуется

условием $|\underline{H}_k - \underline{M}_s| > H_s$. Тогда при приложении внешнего поля пленка может быть только переключена между двумя насыщенными состояниями. В таком случае пленка демонстрирует настоящий эффект двончной памяти, как в случае памяти на магнитных сердечниках.

Замечено, что в используемых в экспериментах пленках с большими площадями при приложении полей, больших $|\underline{H}_k - \underline{M}_s|$, намагниченность не переключается сразу из состояния с одним направлением намагниченности в другое. Это обусловлено некоторым количеством дефектов, всегда обнаруживаемых в пленках (обычно 3—5 на см^2) и на границе подложки. Когда приложено внешнее поле требуемой величины, переворот намагниченности начинается на этих дефектах уже при меньших величинах поля. Эти дефекты являются центрами образования доменов. Измеренная петля гистерезиса, следовательно, не является столь крутой и прямоугольной, как предполагалось до сих пор теоретически. С другой стороны,

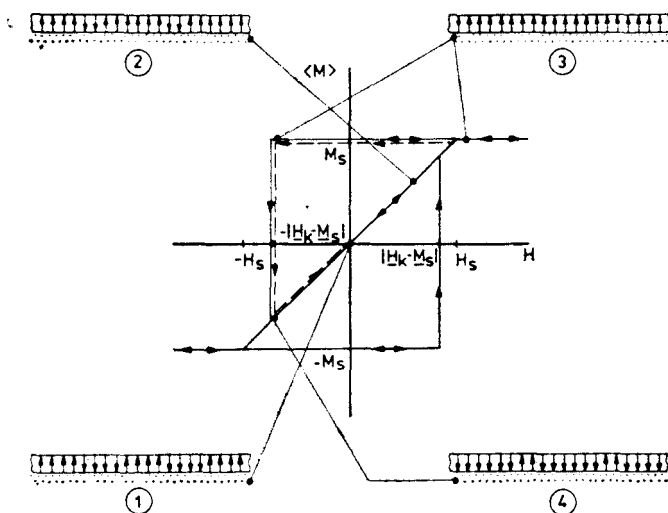


Рис. 1.8. Петля гистерезиса для $|\underline{H}_k| > |\underline{M}_s|$ и $|\underline{H}_k - \underline{M}_s| < H_s$. Исходя из размагниченного состояния пленки (1), средняя намагниченность $\langle M \rangle$ увеличивается вместе с $|H|$ (2). При $|H| = H_s$ пленка является насыщенной ($\langle M \rangle = |\underline{M}_s|$), (3). Это состояние остается устойчивым, если $|H|$ уменьшено до нуля и изменено свое направление на обратное потому, что поле анизотропии \underline{H}_k превышает размагничивающее поле $H_N = -\underline{M}_s$. Если величина обратного по направлению поля превышает $|\underline{H}_k - \underline{M}_s|$, намагниченность становится неустойчивой и снова подразделяется на доменную структуру (4). Выключение приложенного поля снова приводит пленку в состояние (1) (пунктирная линия). С другой стороны, увеличение H до значения $-H_s$ насыщает пленку в противоположном направлении,

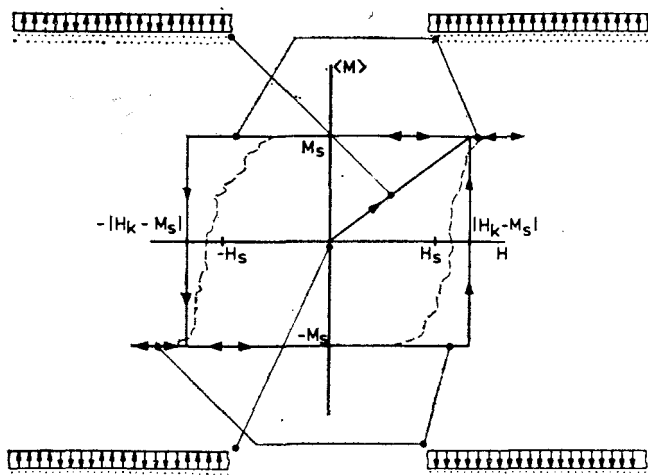


Рис. 1.9. Петля гистерезиса для $|H_k| > |M_s|$ и $|H_k - M_s| > H_s$. В случае насыщения пленки внешним полем $|H_k| \leq H_s$ изменение направления намагниченности на обратное может быть достигнуто, только если перевернуто направление приложенного поля и его величина превышает $|H_k - M_s|$. Используемые в экспериментах пленки с дефектами, служащими центрами образования доменов, принципиально дают гистерезисные петли вида, показанного пунктиром.

переключение доменов посредством приложения внешнего поля является полезным методом определения дефектов в пленках большой площади [10].

1.4. Оптические свойства магнитооптических гранатовых пленок

Основные оптические свойства магнитооптических гранатовых пленок описываются эффектом Фарадея и поглощением в материале [6, 12, 15—18].

Эффект Фарадея, как кратко изложено в разд. 1.2, возникает из-за того, что материал имеет различный показатель преломления для света с круговыми поляризациями по часовой и против часовой стрелок. Фарадеевское вращение плоскости поляризации Θ пропорционально оптическому пути d в среде

$$\Theta = \Theta_F \cdot d, \quad (1.1)$$

где удельный угол поворота плоскости поляризации Θ_F , приходящийся на единицу длины, является функцией длины волны, что показано на рис. 1.10 для типичного магнитооптического граната.

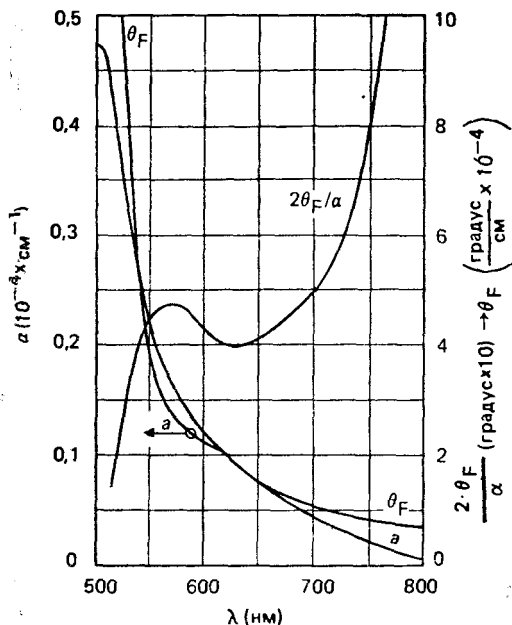


Рис. 1.10. Коэффициент поглощения α , удельный угол поворота плоскости поляризации θ_F и показатель качества материала $2\theta_F/\alpha$ типичной магнитооптической пленки с высокой степенью замещения висмутом.

Кроме того, магнитооптический материал обнаруживает поглощающие свойства, описываемые коэффициентом поглощения α . Видно, что α , так же как и удельный угол поворота плоскости поляризации, возрастает при переходе в голубую область спектра. Как показано на рис. 1.10, оба параметра определяют теоретический коэффициент пропускания и контраст модулирующей системы.

1.4.1. Теоретический коэффициент пропускания монохроматического света

В рамках нижеизложенного полный коэффициент пропускания определяется отношением интенсивности света на выходе анализатора к интенсивности неполяризованного света на входе поляризатора.

Если предположить, что все оптические поверхности в схеме покрыты идеально просветляющими покрытиями, тогда полный коэффициент отражения определяется только двумя состояниями намагниченности и пятью параметрами:

а) коэффициентом поглощения магнитооптического материала, заданным в см^{-1} ;

- б) удельным углом поворота плоскости поляризации Θ_F магнитооптического материала, заданным в градусах/см;
 в) толщиной магнитооптического слоя d ;
 г) углом ϕ_0 между осью анализатора и осью y (рис. 1.3);
 д) коэффициентами пропускания T_P и T_A поляризатора и анализатора ($\approx 100\%$ для кристаллических поляризаторов и обычно $70\text{--}80\%$ для поляризующих пленок).

В зависимости от двух состояний намагниченности полный коэффициент пропускания задается выражениями

$$T_t = 0,5T_P T_A \exp(-\alpha d) \sin^2(\phi_0 + \Theta_F d) \quad (1.2)$$

и

$$T_b = 0,5T_P T_A \exp(-\alpha d) \sin^2(\phi_0 - \Theta_F d). \quad (1.3)$$

Коэффициент пропускания T_t описывает «пропускающее» свет состояние устройства, а T_b — соответственно «запирающее» состояние. Отношение T_t/T_b называется оптическим контрастом переключения C .

$$C = \frac{\sin^2(\phi_0 + \Theta_F d)}{\sin^2(\phi_0 - \Theta_F d)}. \quad (1.4)$$

Следовательно, эта величина определяется только параметром материала $\Theta_F \cdot d$ и углом поворота анализатора ϕ_0 .

Так как параметры материала зависят от длины волны света λ , то T_t , T_b и C имеют вид

$$T_t = T_t(\lambda), \quad T_b = T_b(\lambda), \quad C = C(\lambda). \quad (1.5)$$

При заданной длине волны λ контраст переключения и коэффициент пропускания могут быть оптимизированы в зависимости от толщины d магнитооптической пленки. Это дает $\phi_0 = \Theta_F d$ при $C = \infty$:

$$d_{\text{опт}} = \frac{1}{2\Theta_F} \operatorname{arctg} \left\{ 2 \left(\frac{2\Theta_F}{\alpha} \right) \right\}, \quad (1.6)$$

$$T(\lambda)_{\text{макс}} = 0,5T_A T_P \frac{4(2\Theta_F/\alpha)^2}{1 + 4(2\Theta_F/\alpha)^2} \exp \left\{ -\frac{\operatorname{arctg} \{ 2(2\Theta_F/\alpha) \}}{(2\Theta_F/\alpha)} \right\}. \quad (1.7)$$

Таким образом, максимум $T(\lambda)$ является функцией параметра $2\Theta_F/\alpha$, который соответственно называется показателем качества материала (ПКМ). На рис. 1.10 также показана зависимость ПКМ от длины волны света.

На рис. 1.11 представлены зависимости от ПКМ коэффициента пропускания $T(\lambda)_{\text{макс}}$ и величины фарадеевского вращения плоскости поляризации $2d_{\text{опт}}\Theta_F$, необходимой для достижения максимума пропускания.

В настоящее время ПКМ получаемых в лабораторных условиях материалов составляет в видимой области менее 50° . Ве-

личина ПКМ пропорциональна количеству введенного висмута, однако высокая степень содержания висмута приводит к очень высокой степени анизотропии, что затрудняет процесс переключения (см. разд. 1.5). Заметим, что за счет использования максимально возможной степени замещения висмутом в лабораторных условиях для видимой области спектра удалось

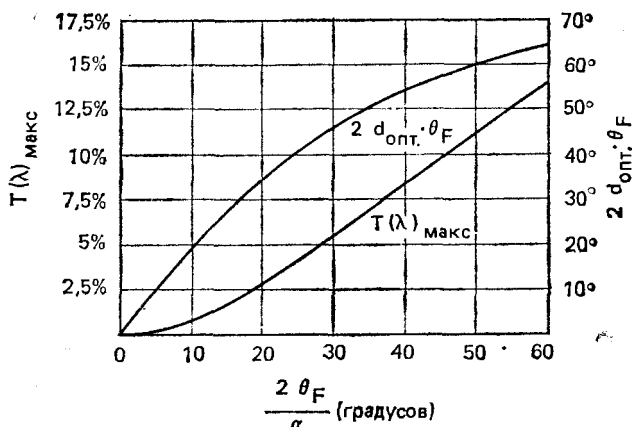


Рис. 1.11. Зависимости максимального значения коэффициента пропускания $T_{t, \text{макс}}$ и оптимальной величины фарадеевского вращения $2d_{\text{опт}} \cdot \theta_F$ от показателя качества материала.

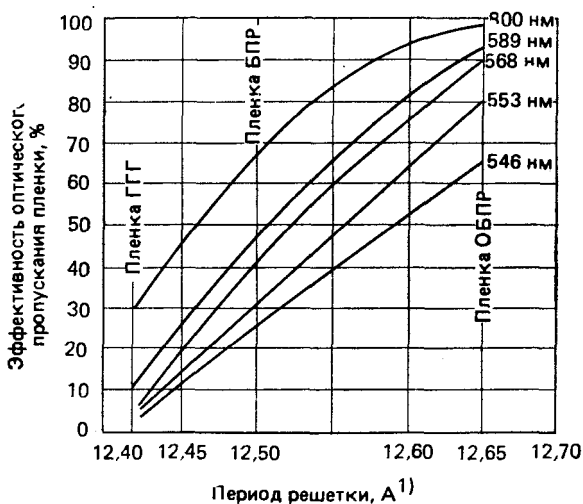


Рис. 1.12. Эффективность пропускания света пленкой $T_{t, \text{лин. поляр}}$ для линейно-поляризованного света на входе устройства в зависимости от периода решетки кристалла. (ГГГ — гадолиниво-галлиевый гранат; БПР — материал с большим периодом решетки; ОБПР — материал с очень большим периодом решетки.)

достичь величины ПКМ, равной 60° . Однако эти материалы еще далеки от совершенства и содержат некоторое количество свинца, увеличивающего поглощение. Теоретически наивысший ожидаемый коэффициент пропускания этого класса материалов может составлять 13% [6, 13, 17].

При использовании лазерного излучения, уже являющегося линейно поляризованным, коэффициент пропускания будет вдвое выше того, который указан на рис. 1.11: $T_{t, \text{лин. поляр}} = 2T_t$.

В лабораторных условиях для новейших низкопоглощающих, не содержащих свинца пленок с просветляющими покрытиями, при работе с полупроводниковыми инжекционными лазерами на длине волны $\lambda = 800$ нм были получены очень высокие коэффициенты пропускания $T_{t, \text{лин. поляр}}$ — более 65% при оптимизированной толщине (ПКМ при этом составлял 230°). Эти пленки относят к классу пленок с большим периодом решетки (БПР), в которых период решетки увеличивается пропорционально увеличению содержания висмута. Возможны методы дальнейшего увеличения периода решетки и увеличения содержания висмута, приводящие к дальнейшему увеличению $T_{t, \text{лин. поляр}}$ до значений, превышающих 95% в ближней ИК области лазерного излучения, 90% для линии спектра Na в 589 нм и 60% для зеленой линии в 546 нм. Эти планируемые к использованию составы пленок относят к классу материалов с очень большим периодом решетки (ОБПР) (рис. 1.12) [19]. Последние работы по ионной имплантации открыли пути к увеличению анизотропии, так что можно надеяться, что низкие переключающие поля все же будут достигнуты [20].

1.4.2. Спектральная зависимость коэффициента пропускания

При освещении магнитооптического модулятора белым поляризованным светом следует принимать во внимание дисперсию фарадеевского вращения плоскости поляризации и коэффициента поглощения. На рис. 1.13 представлены зависимости коэффициента пропускания от длины волны в «блокирующем» и «пропускающем» режимах для ряда типичных значений угла установки анализатора ϕ_0 . При смещении в голубую область спектра свет полностью блокируется из-за высокого поглощения. Полная блокировка света с длинами волн более $\lambda = 500$ нм достигается только для определенных, зависящих от угла ϕ_0 длин волн (например, $\lambda = 545$ нм при $\phi_0 = 25^\circ$).

Наряду с «блокированием» длины волны в данном случае имеет место снижение контраста переключения. Чем больше смещение в красную и инфракрасную области спектра, тем разница между двумя переключающими состояниями становится

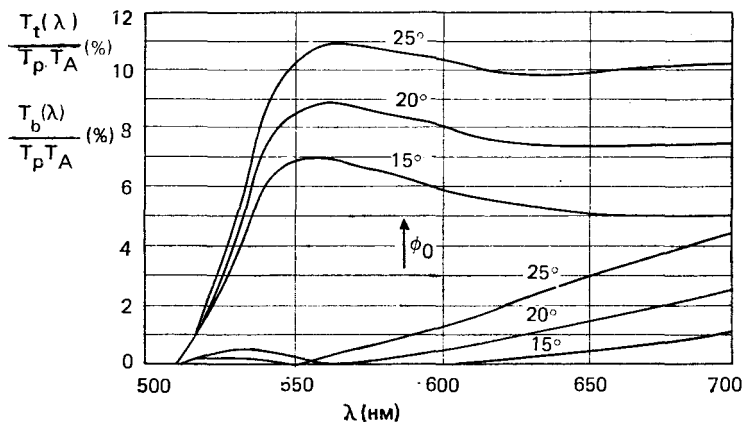


Рис. 1.13. Спектральная зависимость коэффициента пропускания света при различных углах установки анализатора в случае неполяризованного света на входе. Толщина слоя магнитооптической пленки $d=5$ мкм.

все меньше и меньше. Чтобы дать представление о средней величине контраста, достигаемого при использовании модулятора для модуляции света с предельной длиной волны λ_b , можно привести зависимость среднего контраста, которая определяется выражением

$$C_p = \frac{\int_{500 \text{ нм}}^{\lambda_b} T_t(\lambda) d\lambda}{\int_{500 \text{ нм}}^{\lambda_b} T_b(\lambda) d\lambda} \quad (1.8)$$

и представлена на рис. 1.14. Имеются типичные максимальные значения контраста, зависящие от угла установки анализатора и предельного значения λ_b . Максимально достижимый контраст уменьшается при увеличении предельного значения длины волны [16].

Для оптимизации модулятора в заданном оптическом диапазоне необходимо также учитывать толщину слоя d . На рис. 1.15 величина среднего контраста представлена как функция толщины слоя для различных значений ϕ_0 и λ_b . Видно, что при оптимальной толщине пленки значение среднего контраста колеблется в диапазоне от 21 до 23. На рис. 1.15 также представлена спектральная зависимость коэффициента пропускания, проинтегрированная по спектральному диапазону от 500 нм до λ_b .

$$P^* = \int_{500 \text{ нм}}^{\lambda_b} T_t(\lambda) d\lambda. \quad (1.9)$$

Данная величина соответствует световой мощности, получаемой на выходе устройства. Световая мощность растет при увеличении толщины слоя, но не безгранично — существуют огра-

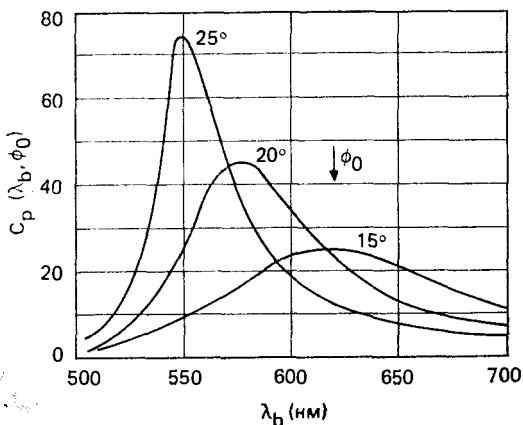


Рис. 1.14. Зависимость средней величины контраста C_p от предельного значения длины волны λ_b при различных углах установки анализатора (для параметров материала, указанных на рис. 1.10).

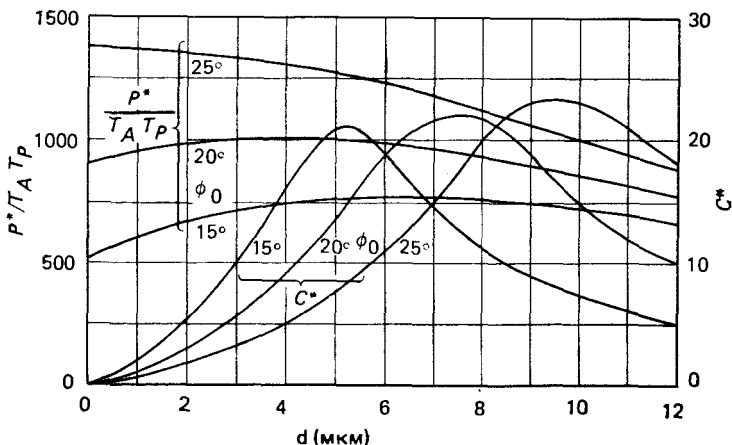


Рис. 1.15. Зависимость средней величины контраста C^* и интегрального коэффициента пропускания $P^*/T_A T_P$ от толщины магнитооптической пленки d для тех же параметров материала, что и на рис. 1.10, и для предельной длины волны $\lambda_b = 650 \text{ нм}$.

ничения, накладываемые технологией изготовления. В настоящее время трудно достичь образования доменных структур в магнитооптической пленке при толщине более 5 мкм. Однако в будущем путем усовершенствования технологии могут быть изысканы некоторые возможности для увеличения выходной мощности [16].

1.4.3. Практические ограничения полной эффективности

Теоретические значения эффективности, обсуждавшиеся до сих пор, не могут быть достигнуты в реальных устройствах, так как в различных компонентах устройства имеются дополнительные механизмы потерь:

1. Канавки между магнитооптическими ячейками и части ячеек, покрытые металлическими проводниками с целью адресации, не пропускают свет. Таким образом, в экспериментальных устройствах потери света за счет маскирования составляют от 20 до 30%.

2. Большинство поляризующих элементов поглощают свет. Этим явлением можно пренебречь только для очень дорогих кристаллических поляризаторов. Дешевые пластиковые поляризаторы вносят дополнительные потери, составляющие по меньшей мере около 30%.

3. В основной оптической схеме модулятора имеется семь оптических поверхностей, вызывающих потери из-за отражения. На шести из них могут быть нанесены просветляющие покрытия. Обычно с помощью многослойных покрытий потери могут быть снижены до 0,2—0,5% на одну поверхность. Так как имеется различие показателей преломления между магнитооптическим слоем и подложкой ($n_{\text{подложки}} \approx 1,9$, $n_{\text{пленки}} \approx 2,3$), то в магнитооптическом устройстве имеются внутренние потери на отражение, составляющие в настоящее время около 0,9%. Общие потери на отражение составляют около 2—3% (при наличии просветляющих покрытий).

Практическая общая эффективность пропускания света из-за перечисленных источников потерь будет уменьшена на 44—48% по сравнению с теоретическим коэффициентом пропускания. Следовательно, средняя общая эффективность пропускания света составляет 5% (для материалов, изготовлявшихся в прошлом). При этом считается, что наилучшее теоретическое значение коэффициента пропускания в видимой части спектра составляет 10—11%. Эти параметры могут быть улучшены за счет использования новых высокоэффективных оптических пленок (см. 1.4.1).

1.4.4. Ограничения величины контраста

Как кратко описано в разд. 1.4.2, контраст оптического переключения при освещении в широком диапазоне длин волн в основном ограничен дисперсией материалов. В узкой полосе частот либо при монохроматическом свете величина контраста в основном определяется эффектом Фарадея. Однако имеется дополнительный эффект, встречающийся в материалах и получивший название кругового дихроизма, вследствие которого коэффициент поглощения для правой и левой поляризаций проходящего через пленку света с круговой поляризацией несколько различается. Отсюда складываемые компоненты света вследствие эффекта Фарадея имеют несколько различающиеся амплитуды и в сумме дают частично эллиптически поляризованный свет. Эллиптически поляризованный свет не может быть полностью блокирован анализатором. Эффект кругового дихроизма достаточно сильно проявляется в голубой и ослабевает по мере смещения в красную область спектра. В голубой области этот эффект приводит к тому, что значение контраста оказывается менее 10, в зеленой он приблизительно равен 50, а в красной этим эффектом можно пренебречь [17].

Эллиптичность может быть скомпенсирована надлежащим подбором оптической фазовой пластины, являющейся дополнительным компонентом модулирующего устройства. В этом случае может быть снова получен очень высокий контраст. Для модуляции монохроматического лазерного излучения с помощью методики компенсации может быть достигнута величина контраста более 1000.

Третий эффект, ограничивающий оптический контраст, происходит вследствие внутреннего отражения на границе между магнитооптической пленкой и подложкой. Если поверхность подложки не покрыта просветляющим покрытием, часть модулированного света отражается обратно. На внутренней границе раздела происходит второе отражение, так что компонента света снова направляется на выход из устройства. Наконец, эта часть проходит через пленку в третий раз, приводя к тому, что полное фарадеевское вращение составляет величину в 3 раза большую, чем для основной доли модулированного света, т. е. $3\Theta_{Fd}$ вместо Θ_{Fd} . Вследствие того что эта компонента имеет иную плоскость поляризации, она не блокируется анализатором и, таким образом, снижает контраст. Более точно выражаясь, возникают многократные отражения этого типа. Однако вследствие поглощения в материале последующими отражениями можно пренебречь. Для угла установки поляризатора, соответствующего условию $\phi_0 = \Theta_{Fd}$ ($C = \infty$ в отсутствие дихроизма и отражения), окончательное значение контраста приближенно описывается формулой

$$C_{\text{отраж.}} = \frac{\sin^2(2\theta Fd)}{\sin^2(4\theta Fd) e^{-2\alpha d} R_{\text{поверхности}} R_{\text{внутренний}}}, \quad (1.10)$$

где $R_{\text{поверхности}}$ и $R_{\text{внутренний}}$ — коэффициенты отражения на поверхности магнитооптической пленки и соответственно на границе раздела между пленкой и подложкой.

Для параметров, заданных на рис. 1.10, и показателей преломления пленки и подложки, соответственно равных 2,3 и 1,9, значение контраста, ограниченное отражением на длине волны, например $\lambda = 600$ нм, составит 72. При смещении в зеленую область спектра эффект уменьшается, так как растут потери на поглощение.

Эффекты многократного отражения в пленке, в случае когда требуется модулировать когерентный свет, должны анализироваться самым серьезным образом. Они могут послужить причиной возникновения сильных интерференционных эффектов, если на поверхность магнитооптической пленки со всей тщательностью не будет нанесено просветляющее покрытие.

1.5. Конструкция магнитооптических модуляторов

1.5.1. Создание структур магнитооптической пленки

Пространственный модулятор состоит из матрицы отдельных, переключающих свет ячеек. В магнитооптических модуляторах, в частности, переключение достигается путем обращения направления намагниченности магнитных доменов в устойчивой структуре доменов. Однако, как указано в общих чертах в разд. 1.3, устойчивая структура доменов не может быть создана в однородных магнитооптических пленках вследствие высокой подвижности магнитной доменной стенки. Следовательно, пленка должна состоять из отдельных островков магнитооптического материала, разделенных между собой канавками. Пример структурированной пленки показан на рис. 1.2. Каждый островок является основой переключающей ячейки.

Из соображений, изложенных в разд. 1.3, вытекает, что магнитооптический материал должен быть составлен таким образом, чтобы поле анизотропии $|H_k|$ превышало намагниченность насыщения $|M_s|$, и разность $|H_k - M_s|$ была бы много больше поля насыщения H_s . Тогда намагниченность в каждой ячейке останется в устойчивом и насыщенном состоянии (как только она насыщена), т. е. каждый островок будет сохранять устойчивый магнитный домен даже тогда, когда приложено внешнее поле противоположного направления, имеющее амплитуду до $|H_k - M_s|$ (см. рис. 1.9). При этих условиях в магнитооптических модуляторах создается устойчивая структура доменов.

Для переключения намагниченности в отдельной ячейке, описываемой до сих пор, следует прикладывать локальные магнитные поля, превышающие $|\underline{H}_k - \underline{M}_s|$. Так как поле анизотропии пленки с добавками висмута является очень большим (см., например, рис. 1.16), локальное переключение становится достаточно трудным, особенно если при использовании интегральной технологии переключающее магнитное поле надо подвести с помощью металлического тонкопленочного проводника.

Для реализации переключения при более низких переключающих полях были разработаны две различные методики, причем и в той, и в другой поле анизотропии локально уменьшено в пределах каждой ячейки. Это достигается путем локального нагрева ячейки в первой методике и локальной ионной имплантацией во второй.

1.5.2. Модуляторы с термомагнитным переключением

1.5.2.1. Принцип переключения

Намагниченность как сумма дипольных моментов ионов является чувствительной к температуре и падает при приближении к точке Кюри. В ферромагнитных гранатах суммарная намагниченность представляет собой небольшую по величине разность сильных антипараллельных намагниченностей, возникающих вследствие различных дипольных моментов ионов с различными чувствительностями к температуре. В итоге суммарная намагниченность достаточно сильно зависит от температуры.

При термомагнитном переключении используется чувствительность намагниченности в области больших температур. При определенной температуре, называемой температурой компенсации (рис. 1.16), правильным подбором состава граната даже возможно получить нулевую суммарную намагниченность.

При температурах, больших температуры компенсации для указанного материала, максимальная намагниченность насыщения, как правило, очень мала по сравнению с полем анизотропии. Поле анизотропии в свою очередь при температуре компенсации становится бесконечно большим (оно определено как сила, действующая на намагниченность) и спадает, когда температура повышается (рис. 1.16). Этот эффект используется на практике для уменьшения разности $|\underline{H}_k - \underline{M}_s|$ путем локального нагрева магнитооптического материала (термомагнитное переключение).

В используемых на практике материалах температура компенсации лежит в пределах от -10 до 0°C . При нагреве приблизительно до 200°C поле анизотропии, уменьшенное на ве-

личину намагниченности насыщения $|\underline{H}_k - \underline{M}_s|$, очень низко. При этом становится возможным переключение в полях порядка 5 кА/м. В результате появляется домен с противоположным направлением ориентации намагниченности, зародившийся в пределах нагретого участка. Если приложенное переключающее поле в то же время превышает поле насыщения H_s ненагретой части ячейки, новый домен увеличивается за счет движения доменной стенки (в соответствии с линией, проходящей на рис. 1.9 от центра в направлении H_s) до тех пор, пока наконец намагниченность ячейки полностью не достигнет нового насыщенного состояния с противоположной намагниченностью. Таким образом, во время переключающего цикла имеют место два процесса. Во-первых, при подаче поля к локально нагретому участку зарождается домен и, во-вторых, осуществляется переход ячейки в насыщенное состояние за счет перемещения доменной стенки под действием силы, порождаемой приложенным полем. Если поле выключить на промежуточной стадии, ячейка может остаться в размагниченном состоянии с двумя доменами.

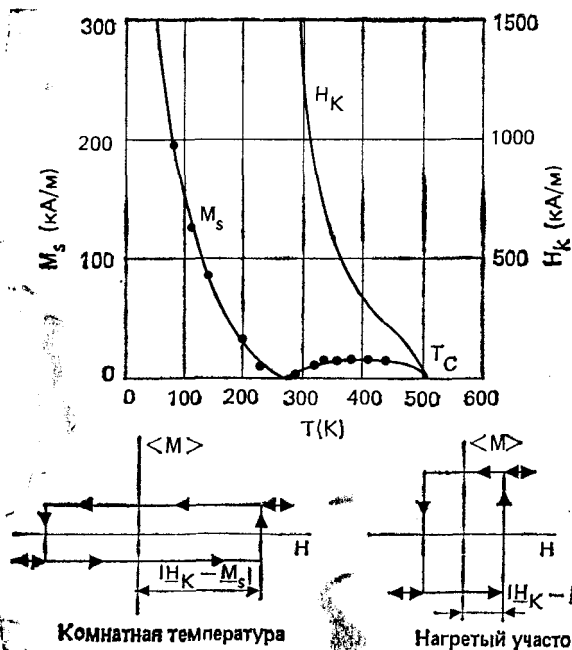


Рис. 1.16. Зависимости намагниченности насыщения и поля анизотропии от температуры для магнитооптического материала, обладающего точкой компенсации. Внизу показана типичная петля гистерезиса при комнатной температуре и соответствующей нагретому участку повышенной температуре.

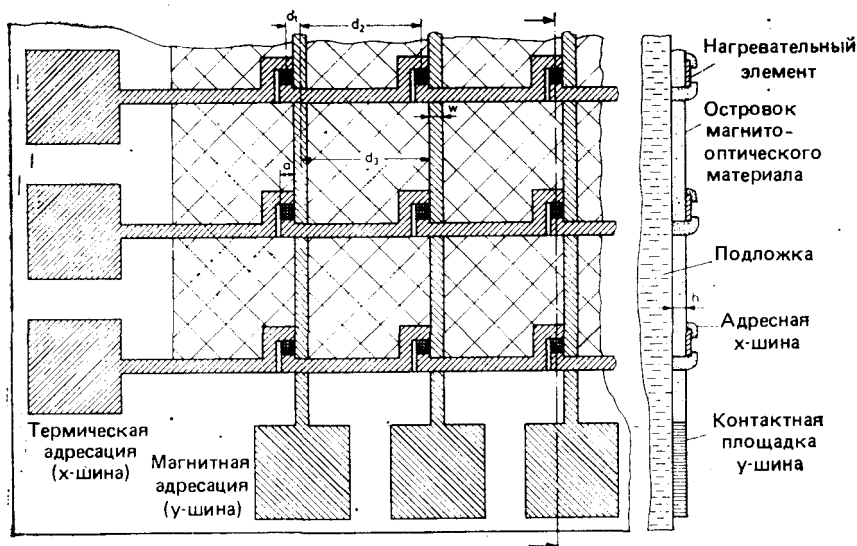


Рис. 1.17. Схема магнитооптического модулятора с термической адресацией. Нагревающий элемент перекрывает часть площади магнитооптической ячейки. Нагревающие элементы в одном ряду соединены, и адресация осуществляется с помощью x-шины. Адресация в столбцах осуществляется магнитными y-шинами. На пересечении x- и y-шины разделены изолирующим слоем, не показанным на схеме.

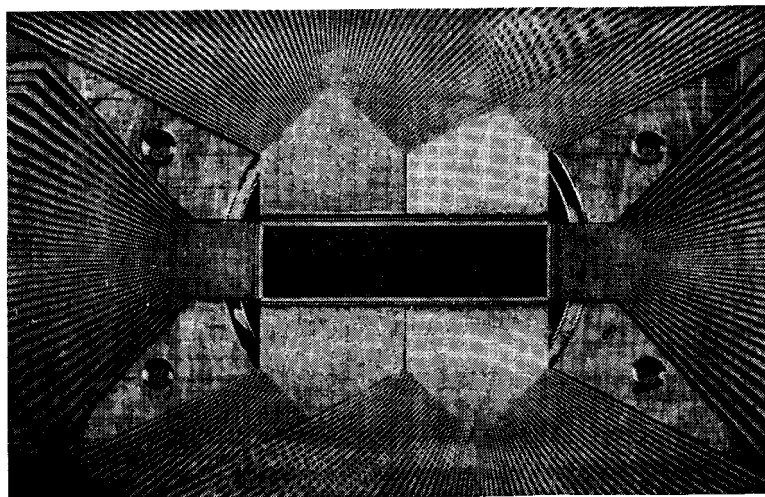


Рис. 1.18. Модулятор 256×64 элементов с разветвленной системой шин для соединения с управляющей электронной схемой [21]. Размеры модулятора — 32×8 мм.

Экспериментальная реализация этой концепции схематично показана на рис. 1.17 и 1.18 [8, 21]. Тонкопленочное сопротивление нанесено на край переключающей ячейки и соединено с тонкопленочной металлической адресной шиной. При пропускании токового импульса выделяется омическое тепло, и материал снизу под резистором нагревается. Магнитное поле для переключения создается второй металлической шиной, располагаемой перпендикулярно первой.

Типичная переключающая характеристика показана на рис. 1.19. Нагрев до достаточно высокой температуры требует около 15 мкс. После чего импульс магнитного поля переключает ячейку за время ≤ 10 мкс (сумма интервалов времени, требуемых на зарождение доменов и перемещение магнитной доменной стенки по ячейке). Следовательно, в пределе могут быть получены 40 000 тепловых циклов в секунду.

1.5.2.2. Проблемы перекрестной адресации [22]

В магнитооптическом модуляторе с перекрестной y -адресацией нагревающие резисторы вдоль одного ряда ячеек соединены

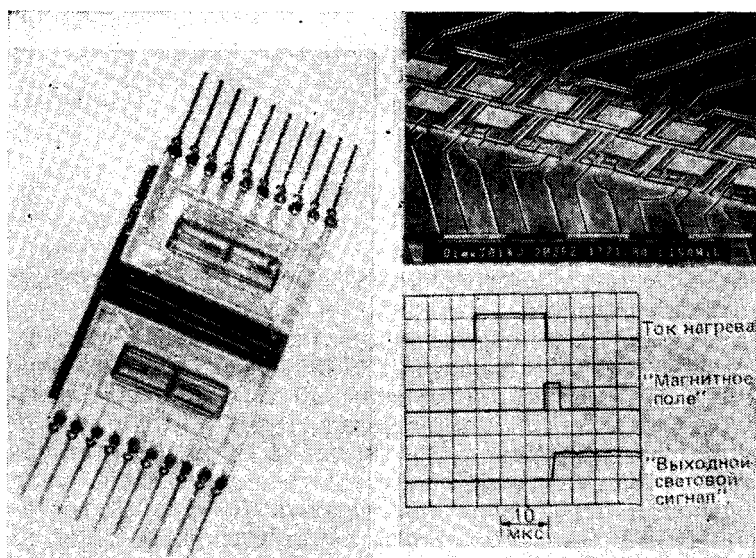


Рис. 1.19. Матрица переключения линейно-поляризованного света с 512 объединенными на одном чипе переключающими ячейками, расположенными с шагом в 62,5 мкм. На изображении тонкопленочной структуры, полученном с помощью электронного микроскопа, видны зигзагообразно расположенные ячейки, нагревающие элементы и тонкопленочная адресная шина. Справа внизу показаны типичные переключающие характеристики.

последовательно. В перпендикулярных им канавках размещены шины, по которым подаются импульсы магнитного поля. Чтобы добиться устойчивого переключения, резисторы должны быть близко расположены к управляющим магнитным шинам. Переключающее поле задается выражением

$$H_1 = \frac{I}{2\pi d_1}, \quad (1.11)$$

где I — протекающий в магнитной управляющей шине ток, а d_1 — расстояние от центра управляющей шины до резистора (см. рис. 1.17 и 1.18). Переключение определяется условием $|\underline{H}_1| > |\underline{H}_k - \underline{M}_s|_{\text{нагретого материала}}$. Поле на наиболее удаленном крае от управляющей шины равно $H_2 = I/2\pi d_2$. Это поле должно превышать поле насыщения H_s «холодного» материала, т. е. должно выполняться неравенство $H_2 > H_s$. Чтобы избежать переключения соседних ячеек, поле

$$H_3 = \frac{I}{2\pi d_3} \quad (1.12)$$

должно быть меньше $|\underline{H}_k - \underline{M}_s|$, где d_3 — расстояние от центра магнитной управляющей шины до резистора в соседнем столбце ячеек. Таким образом, ток I ограничен условием

$$\frac{I}{2\pi d_3} < |\underline{H}_k - \underline{M}_s|_{\text{нагретого}} < \frac{I}{2\pi d_1} \quad (1.13)$$

или

$$2\pi d_1 |\underline{H}_k - \underline{M}_s| < I < 2\pi d_3 |\underline{H}_k - \underline{M}_s|. \quad (1.14)$$

Это условие определяет точность, с которой должны быть изготовлены резисторы.

Примером типичной экспериментальной разработки данного типа является матрица 100×100 ячеек, расположенных с шагом в 100 мкм и с канавками в 10 мкм. Площадь резистивного нагревателя при этом должна быть 10×10 мкм². При сравнительно свободном размещении ячеек d_1 составляет 15 мкм, а ток, который пропускают по магнитной управляющей шине и определяют из условия $H_1 = I/2\pi d_1$, составляет около 470 мА (считается, что переключающие поля составляют 5 кА/м). Расстояние до следующего нагревателя составляет около 90 мкм. Это дает фактор защищенности от переключения соседних ячеек в $90/15 = 6$.

Для перемещения образовавшейся магнитной доменной стенки из нагретой области через всю ячейку материал следует подбирать таким, чтобы выполнялось условие $H_s < I/2\pi d_3 \approx 1$ кА/м.

1.5.2.3. Проблемы теплоотвода и электромиграция

В реальных устройствах нельзя пренебрегать омическими потерями в магнитных управляющих шинах. Рассматривая переключающее поле H на расстоянии $d_1 = \frac{1}{2}(\omega + a)$ от центра шины, получим, что омические потери на расстоянии l от шины определяются выражением

$$P_v = H^2 \pi^2 \rho \frac{(\omega + a)^2}{\omega h} l, \quad (1.15)$$

где ρ — удельное сопротивление используемого металла, h — высота шины, а ω — ширина проводника. В случае использования золотой шины при сечении y -шины 10×5 мкм² (максимум h задается толщиной магнитооптической пленки, составляющей около 5 мкм, если шина целиком заполняет канавку) потери, приходящиеся на 1 см, составят около 1 Вт. Эта мощность переходит в тепло за время (менее 10 мкс), в течение которого прикладывается магнитное поле. Энергетические потери за акт переключения составляют довольно высокую величину — 10 мкВт/см. При переключении матричного модулятора эту величину следует умножить на число переключающих ячеек. Например, переключение 10^4 ячеек в матрице размером 1×1 см дает величину выделяемой мощности в 100 мВт/см за время переключения матрицы и соответственно среднюю мощность в 100 мВт за 1 с.

В соответствии с формулой можно ожидать небольшого снижения потерь в случае уменьшения ширины проводника ω . Однако, поскольку размеры нагревательного элемента не могут быть резко уменьшены из-за возникающей в этом случае проблемы воспроизводимости устройства, функция $(\omega + a)^2 / \omega h$ при уменьшении ω резко не изменится. Любого значительного снижения тепловых потерь можно ожидать только за счет улучшения свойств материалов (т. е. уменьшения $|\underline{H}_k - \underline{M}_s|$), так как $P_v \sim H^2$. Второй эффект, который следует рассмотреть, это электромиграция в тонких проводниках при высоких плотностях тока [23]. Плотности тока, превышающие обычно 10^6 А/см², вызовут химическое разрушение материала проводника [24].

Для обсуждавшихся до сих пор численных примеров плотность тока ограничивается этим пределом. Дальнейшее снижение размера проводников при том же токе, следовательно, невозможно.

Тепловая энергия, необходимая для переключения магнитооптической ячейки, составляет около 0.5 мкДж на ячейку. При типичном времени нагрева ячейки в 15 мкс это дает пиковую приходящуюся на 1 ячейку мощность в 33 мВт. В матри-

це последовательно соединены в одном ряду N ячеек, что дает $N \times 33$ мВт на 1 ряд (3,3 Вт в случае матрицы в 100×100 элементов).

Так как за тепловым импульсом должны следовать импульсы магнитного поля длительностью 10 мкс, минимальное время цикла переключения составляет 25 мкс, что соответствует 40 000 циклам в секунду. При переключениях с этой предельной скоростью средняя мощность, рассеиваемая в нагревательных элементах, составляет $N \times 33$ мВт $\times 15/25 = N \times 20$ мВт (2 Вт для представленного численного примера). Управляющее напряжение, прикладываемое к x -шине, следует выбирать максимально возможным, чтобы сохранить на низком уровне ток и дополнительные потери в x -шине. Напряженне в 20 В, например, приводит к управляющим токам в 100 мА и дополнительным потерям мощности около $1/20$ Вт, если поперечное сечение x -шины также составляет 10×5 мкм².

Пиковая мощность, необходимая для проведения пошаговой адресации, таким образом, составляет около 3 Вт для матрицы в 100×100 элементов. Для переключения с максимальной скоростью 40 000 переключений в секунду требуется средняя мощность в 3 Вт. Дальнейшее увеличение может произойти при построчном переключении ячеек. Одновременное переключение M ячеек требует в M раз большей мощности на y -шине (например, для 10 ячеек — 10 Вт и т. д.). Переключение одновременно всех ячеек матрицы потребовало бы слишком большой пиковой мощности, и, кроме того, возникают жесткие проблемы перекрестных наводок. В предельном случае магнитные поля смежных шин-столбцов могут в сумме давать те же порядки величин, что и адресующие поля в ячейках. При одновременном переключении ячеек следует выбирать их на максимальном расстоянии друг от друга.

1.5.2.4. Линейки элементов с термической адресацией

Для линейной матрицы элементов, состоящей только из одного столбца ячеек, термомагнитное переключение, осуществляемое с помощью нагревательных элементов, является наиболее подходящим, так как управляющие токи в нагревательном элементе могут составлять только несколько миллиампер при напряжениях около 8 В. Это позволяет осуществлять непосредственное и параллельное управление с помощью полупроводниковых элементов с высокой степенью интеграции. Избирательное переключение в этом случае достигается с помощью термической адресации подачей магнитного поля сразу на всю матрицу. Этот принцип используется в магнитооптической матрице переключения света, показанной на рис. 1.19. Управляющие токи для термического переключения поступа-

ют в n -канальных МОП интегральных схем, установленных на каждой из сторон матрицы. При этом может выполняться до 2000 циклов переключения в секунду, что соответствует скорости передачи данных изображения в 1 Мбит/с [25].

1.5.3. Модулятор света «лайтмод» с магнитной адресацией

1.5.3.1. Принцип работы

В устройстве «лайтмод» [11] уменьшение переключающего поля достигается с помощью имплантации отдельных участков магнитооптического материала ионами V^+ (рис. 1.20). Имплантация ионов вызывает внутренние напряжения, направленные перпендикулярно плоскости пленки и параллельно оси поля анизотропии. Эти напряжения снижают поле анизотропии. На рис. 1.21 показано типичное уменьшение величины переключающего поля в зависимости от индуцированных напряжений.

Ионно-имплантированная область переключающей ячейки в «лайтмоде», имеющем пониженные переключающие поля, расположена вблизи точки пересечения x - и y -адресных шин (рис. 1.20). При подаче токовых импульсов I по адресным шинам поле в центре имплантированной области определяется как

$$H_1 = \frac{I}{\pi d_1}, \quad (1.16)$$

где d_1 — расстояние от центра адресной шины до центра имплантированной области. Для образования домена с обратным направлением намагниченности это поле по величине должно по меньшей мере превосходить переключающее поле. И если приложенное поле оказывается достаточно большим, чтобы продвинуть доменную стенку по ячейке, то новый возникший домен за счет движения стенок начинает увеличиваться в размерах. И опять-таки ячейка переходит в насыщенное состояние, если приложенное поле превышает поле насыщения по всей площади ячейки. Чтобы избежать переключения соседней ячейки по отношению к адресуемой, ток, пропускаемый по шине выбора ячейки, не должен превышать в два раза минимальное значение, необходимое для переключения в адресуемой точке пересечения шин. Условием селективного переключения, таким образом, является

$$H_{\min} < \frac{I}{\pi d_1} < 2H_{\min}. \quad (1.17)$$

В реальных устройствах следует принимать во внимание статистический разброс переключающего поля в имплантированной области.

Глубина имплантации ионов составляет величину, близкую

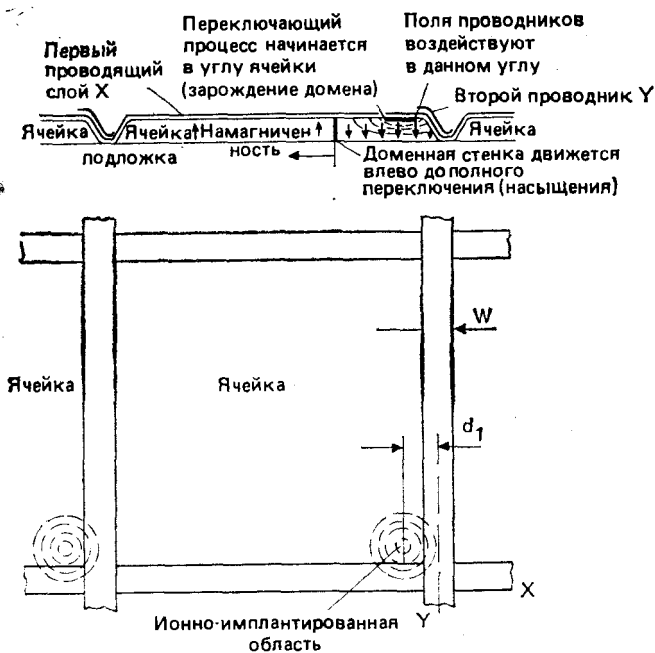


Рис. 1.20. Конструкция «лайтмод» с x - и y -адресующими шинами и ионно-имплантированными областями, расположенными вблизи от точек пересечения x - и y -шин.

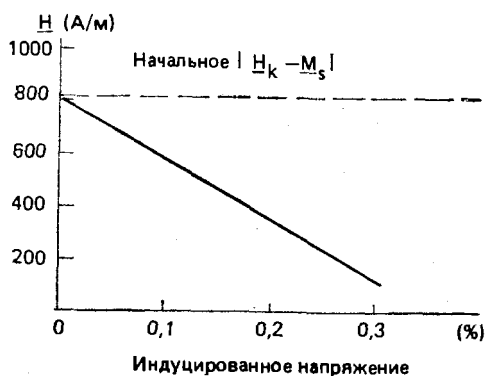


Рис. 1.21. Снижение переключающего поля с помощью имплантации ионов V^+ и отжига (при 450°C в течение $1/2$ ч в атмосфере N_2) в зависимости от величины индуцированных напряжений.

0,5 мкм. Детальный анализ процесса зарождения доменов и величины поля, требуемого для продвижения возникшей доменной стенки из имплантированной в неимплантированную область ячейки, показывает, что минимальные переключающие поля H_{\min} должны составлять приблизительно 12 кА/м [10]. Из-за сильного градиента поля анизотропии в центре и на границе имплантированной области возникает градиент энергии магнитной доменной стенки. Это приводит к появлению силы, оказывающей «давление» на стенку, направленной от центра имплантированной области и заставляющей домен снова сжиматься. Чтобы преодолеть это сжатие, требуется минимальное переключающее поле H_{\min} , даже если величина $[H_k - M_s]$ уменьшена до нуля за счет высокой дозы имплантации в центре имплантированной области. Время переключения в основном определяется скоростью движения доменной стенки. Наблюдавшиеся времена переключения находятся в пределах 1 мкс, что соответствует скорости движения стенки около 100 м/с [10]. Переключение, таким образом, происходит намного быстрее, чем у термически адресуемых модуляторов, обсуждавшихся в подразд. 1.5.1, где время переключения ограничено тепловыми процессами. Устройство «лайтмод» с 256×256 элементами, смонтированное на керамической подложке с межсоединениями, показано на рис. 1.2.2.

1.5.3.2. Управляющие токи, проблемы тепловода и электромиграция

Для матрицы в 100×100 ячеек с шагом 100 мкм и шириной канавок 10 мкм минимальные переключающие поля в x - и y -адресных шинах, составляющие $H_{\min} = 12$ кА/м, требуют пропускания импульсов тока величиной около

$$I = H \pi d_1 = 380 \text{ мА.} \quad (1.18)$$

Расстояние d_1 от центра шин до центра имплантированной области в данном случае составляет приблизительно 10 мкм.

Соответствующие омические потери мощности в адресных шинах длиной l тогда составляют

$$P_v = \rho \frac{H^2 \pi^2 d_1^2}{wh} l, \quad (1.19)$$

где wh — поперечное сечение шин (см. рис. 1.20). Толщина адресных шин может быть доведена до значения толщины пленки (обычно составляющая 5 мкм). Для рассмотренного численного примера результирующие потери мощности в золотой шине длиной 1 см составляют около 2,5 Вт. Переключение с максимальной скоростью (1 элемент/мкс) требует подвода мощности около 5 Вт (сумма по x - и y -шинам) к рассматри-

ваемой матрице в 100×100 элементов. В принципе ряды элементов также могут переключаться одновременно за счет одновременного включения y -шин. Однако в этом случае, помимо проблемы создания перекрестных помех смежными шинами с одинаковым направлением тока возникает еще и проблема, связанная с ростом сложности управляющей электронной схемы, что объясняется необходимостью обеспечить подвод пиковой мощности, приходящейся на одну адресную шину (2,5 Вт для обсуждавшегося примера), умноженную на число y -шин. Для простоты в дальнейшем будет предполагаться последовательное переключение элемента за элементом, что приемлемо для большинства применений.

При увеличении размера и числа ячеек потери мощности будут увеличиваться по крайней мере пропорционально размеру. В настоящее время максимальный доступный размер подложек составляет 7,5 см. Следовательно, могут быть реализованы модуляторы, имеющие размер 35×35 мм, что соответствует матрице 350×350 ячеек с шагом 100 мкм. Возможно и уменьшение шага; однако минимальный размер канавок задан толщиной пленки, умноженной на коэффициент, определяемый технологией травления. Так как толщина пленки в большинстве случаев должна быть максимально возможной для достаточного угла поворота плоскости поляризации, снижение размера ячеек снизит потери на оптическое маскирование (до настоящего момента технология травления применялась для пленок толщиной от 5 до 7 мкм и позволяет получить канавки глубиной около 10 мкм).

В примере, приведенном на рис. 1.2, показаны ячейки с минимальным размером в 10 мкм, разделенные канавками в 10 мкм (шаг 20 мкм). Потери на маскирование для этого устройства будут составлять 75% по сравнению с 19% для устройства с шагом 100 мкм и канавками в 10 мкм.

С другой стороны, уменьшение размера ячеек приведет к увеличению скорости переключения, если полагать, что скорость движения магнитной доменной стенки не изменяется. В первом порядке приближения до тех пор, пока рассматривается время магнитного переключения, время переключения матрицы с заданной площадью не зависит от числа адресных шин или размера ячеек в матрице. Однако практически необходимо рассматривать дополнительные времена задержки в электронных управляющих цепях. Они не зависят от размера ячеек и суммируются в зависимости от числа адресных шин. Таким образом, поскольку требуется переключать довольно большие величины импульсных токов, то в конечном счете предельные времена переключения определит управляющая электроника. Заметим также, что, как и в случае модулятора с термомагнитным переключением, здесь следует учитывать

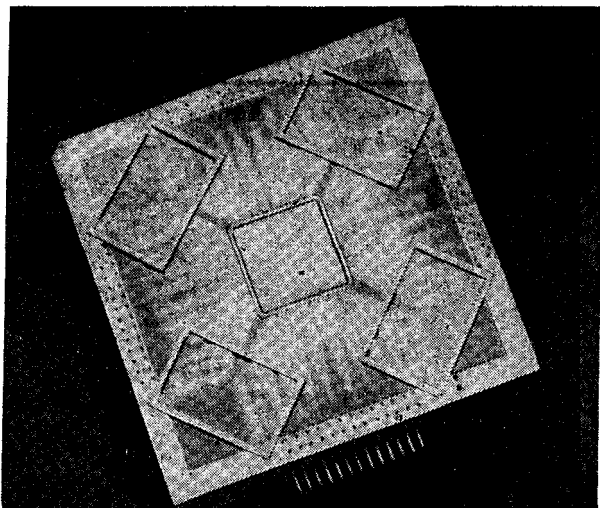


Рис. 1.22. Модулятор «лайтмод» с матрицей в 256×256 элементов, размещенной на керамической подложке схемы. Система межсоединений адресных шин содержит диодную схему выборки для уменьшения числа управляющих шин.

ограничения, связанные с электромиграцией. Для токов адресации в 300—400 мА поперечное сечение адресных шин должно составлять по крайней мере 50—100 $\mu\text{м}^2$, чтобы не подвергнуться действию электромиграции, связанной с разрушением шины при слишком больших плотностях токов.

1.5.4. Способ переключения с помощью магнитного поля смещения¹⁾

В некоторых приложениях может быть полезным использование для переключения дополнительного магнитного поля смещения. Прямоугольная петля гистерезиса и соответствующий ей характер уровней переключения приводят к тому, что переключение полем смещения вызывает снижение импульсов тока, прикладываемых к магнитным адресным шинам, а также снимает остроту проблем, связанных с отводом тепла и управляющей электроникой.

Однако, поскольку магнитное поле смещения должно быть создано довольно большой катушкой, в которой размещается модулятор, оно должно прикладываться только в режиме запись — стирание, предназначенном для работы с изображениями. Высокая индуктивная нагрузка катушки дает миллисе-

¹⁾ В оригинале bias field. — Прим. перев.

кундные времена переключения. Это не удовлетворяет требованиям к времени переключения. При записи изображения в модулятор приложенное поле должно включаться только во время цикла записи. Стирание изображения может быть реализовано без использования магнитных адресных шин в том случае, если обращенное по направлению поле столь же велико, как и минимальное переключающее поле. (Для термомагнитного модулятора тепловые импульсы должны прикладываться только для последовательного стирания, ряд за рядом.) Минимально возможные локальные переключающие поля, создаваемые адресными шинами и накладываемые на поле смещения, определяются статистическим разбросом величины необходимого переключающего поля H в модуляторе. Экспериментальные результаты, полученные для модуляторов с термомагнитным переключением, показали, что до 80 % полного магнитного поля могут быть созданы полем смещения, приводя к уменьшению в 5 раз локальных полей, создаваемых с помощью y -шин.

1.6. Обсуждение результатов и перспектив

Магнитооптическая технология, основанная на монокристаллических гранатах, делает возможной реализацию пространственных модуляторов с большим числом переключающих ячеек при микроскопических размерах ячейки (от 50 до 100 мкм). Максимальная площадь модулятора теперь ограничивается 7,5 см размером существующих гранатовых подложек. Тем не менее на подложке этого размера могли бы быть получены модуляторы с числом элементов 512×512 и шагом 100 мкм или с числом элементов 1024×1024 и шагом 50 мкм.

Переключающие характеристики и характеристики оптического пропускания определяются большим числом параметров, наиболее существенными из которых являются следующие:

1. Магнитные параметры: поле анизотропии, намагниченность насыщения и соответствующее переключающее поле, поле насыщения и подвижность магнитной доменной стенки.

2. Оптические параметры: угол поворота плоскости поляризации, коэффициент поглощения и соответствующий показатель качества материала (ПКМ).

Все эти параметры могут меняться и зависят от состава гранатовой пленки. Широкий диапазон возможных вариантов замещений ионов в решетке граната обеспечивает большое число способов оптимизации материала. Тем не менее до сих пор не было возможным оптимизировать все параметры в одном модуляторе, сделанном из конкретного материала. Например, оптические свойства могут быть улучшены путем замещения висмутом. Однако это одновременно приводит к увеличе-

нию поля анизотропии и переключающих полей. С другой стороны, уменьшение поля анизотропии и переключающих полей до низкого уровня ухудшает оптические свойства.

За последние годы исследований в области материаловедения наметился значительный прогресс в понимании роли различных вариантов замещения и оптимизации состава. Существует реальная надежда на дальнейший успех на пути оптимизации сразу всех параметров.

Главным недостатком существующих в настоящее время устройств является низкий коэффициент пропускания света, определяемый фарадеевским вращением плоскости поляризации, поглощением в материале и толщиной магнитооптической пленки.

Для материалов с высокой степенью замещения висмутом и большими углами поворота плоскости поляризации в существующих устройствах пока не удается достичь оптимальной толщины пленки. Для длины волны в середине видимой области спектра оптимальная толщина составляла бы около 10—12 мкм. Однако современная технология позволяет получить до 5—7 мкм. Таким образом, здесь имеются возможности для улучшения параметров. Фарадеевское вращение плоскости поляризации уменьшается при смещении в красную и инфракрасную области спектра. С другой стороны, имеются «окна» в коэффициенте «поглощения» в инфракрасной области, где ПКМ оказывается на несколько порядков больше, чем в видимой области. Когда для оптической обработки информации используется лазерное излучение в инфракрасной области, имеется возможность создания модуляторов света с намного более высокими оптическими коэффициентами пропускания, чем в видимой области. Трудность, однако, состоит в том, что толщина магнитооптического слоя должна быть намного большей, чем в видимой области, чтобы получить лучший коэффициент пропускания (обычно в интервале до 100 мкм). Это делает технологию изготовления модулятора более сложной. В оптической обработке сигналов может представлять особый интерес получение углов фарадеевского вращения плоскости поляризации в 45 или даже 90°. В видимой области это возможно только при высоком коэффициенте поглощения. Как было показано с помощью простого устройства, в инфракрасной области могут быть получены углы поворота плоскости поляризации в 90° в гранатовых материалах при оптических коэффициентах пропускания до 90% [19, 20, 26]. Однако до сих пор подробно не изучена возможность создания высокоразрешающих пространственных модуляторов в инфракрасной области.

При сравнении термомагнитной адресации и адресации с помощью ионно-имплантированных центров образования доменов в устройстве «лайтмод» становится очевидным, что для

последнего возможны более высокие скорости переключения (около 1 мкс). При этом скорости различаются приблизительно в 25 раз. Пиковые мощности управляющих сигналов в обеих концепциях имеют один порядок величины. Однако средние мощности управляющих сигналов для переключения одного и того же числа ячеек за заданное время намного меньше в «лайтмоде» вследствие больших скоростей и соответственно меньшего времени адресации ячейки.

Вследствие резкого снижения поля анизотропии в зависимости от температуры, характерного для подхода терромагнитного переключения, имеется больше возможностей для оптимизации параметров материала. Таким образом, основанные на терромагнитном переключении реальные модуляторы содержат большое количество висмута и соответственно обладают большей эффективностью модуляции света.

Что касается электронных цепей управления, то в сетях управления магнитных адресных шин, вырабатывающих импульсы тока обеих полярностей в несколько сотен мА, должны использоваться мостиковые схемы. Управляющие каскады шин термической адресации могут быть собраны из простых транзисторных переключателей. Это является преимуществом подхода термической адресации, в частности, если рассматриваются только модуляторы с малым числом шин, но большим числом ячеек на одну шину. Кроме того, малое число нагревающих резисторов позволяет уменьшить управляющие токи до миллиампер, что открывает возможность их использования для управляющих цепей с очень высокой степенью интеграции. Например, линейки элементов, показанные на рис. 1.19, управляются n -канальными МОП-микросхемами с 64 выходными каналами на чип.

Степень качества и плоскостности магнитооптической гранатовой пленки не представляет особых проблем. Слои граната изготавливаются только с тремя-четырьмя дефектами на квадратный сантиметр, и отклонение оптических и магнитных параметров по площади 7,5 сантиметровой пленки составляет только несколько процентов. Статистические отклонения по площади матричного модулятора в основном происходят из-за тонкопленочной технологии изготовления металлических шин, резистивных слоев или ионно-имплантированных областей.

Проблемы «проколов» в масках и сложности травления являются очень жесткими, если требуется изготовить модуляторы с высоким разрешением. В настоящее время выход изготовления модуляторов с большим числом ячеек все еще остается низким.

Размер ячеек обычно может изменяться от 10 до 100 мкм. Так как канавки между ячейками не могут быть сделаны меньше 10 мкм, при толщинах пленок, необходимых для достижения

хорошей эффективности, оптические маскирующие потери уменьшаются вместе с размером ячейки, и относительно большая площадь модулятора оказывается покрыта адресными линиями. Потребляемые пиковые мощности не уменьшаются значительно с размером ячейки. То же самое справедливо и для времени переключения. Метод термической адресации в любом числе требует более 25 мкс, и для «лайтмода» время переключения в субмикросекундной области будет определяться в основном управляющей электроникой из-за необходимости производить переключения высоких пиковых токов.

Автор весьма благодарен м-ру Вильяму Е. Россу (Литтон Дэйта Системс) за предоставление рис. 1.2, 1.12, 1.20 и 1.22, а также профессорам В. Толксдорфу, С. П. Клейтесу, Дж. П. Крумме, П. Хансену, В. Витгеру и К. П. Шмидту из Исследовательской лаборатории компании Филипс, г. Гамбург, за предоставление информации и сведений по термомагнитным модуляторам и свойствам гранатов.

Глава 2

ОПТИЧЕСКИЕ БИСТАБИЛЬНЫЕ УСТРОЙСТВА И ЛОГИЧЕСКИЕ ЭЛЕМЕНТЫ

*Н. Пейгамбарян, Центр оптических исследований,
Университет шт. Аризона, Таксон, шт. Аризона
Дж. Джуэлл, АТТ Белл Лэборэториз, Холмдел,
шт. Нью-Джерси*

2.1. Введение

Изобретение транзистора несколько десятилетий назад вызвало революцию в вычислительной технике и технике связи. Вычислительные машины дотранзисторной эпохи были громоздкими, неудобными и обладали низким быстродействием. Вычислительные машины на транзисторах значительно уменьшились в размерах, а их быстродействие увеличилось; появилась возможность проводить обработку больших массивов цифровой информации. Наряду с этим быстрое развитие технологии электронных вычислительных машин начинает сдерживаться трудностями, возникающими при передаче данных из одной части системы в другую. В отличие от этого оптические методы не предполагают проблем при передаче информации. Вместе с тем уровень развития технологии оптических цифровых устройств значительно отстает от развития электроники. Тем не менее были разработаны оптические бистабильные и логические устройства, аналогичные транзисторам [1]. Транзистор — это устройство, в котором один поток электрических зарядов управляет другим потоком электрических зарядов. Аналогично в оптическом бистабильном устройстве один световой поток управляет другим световым потоком. Бистабильное устройство — это переключатель, который находится в состоянии «включено», если устройство пропускает входящий свет, и «выключено», если устройство блокирует свет. Бистабильные устройства имеют размеры порядка нескольких мкм². Они обладают быстродействием в пикосекундной области и обеспечивают высокую степень параллельности при выполнении операций. Ряд таких бистабильных устройств работает при комнатной температуре и потребляет при этом мощность всего в несколько милливольт.

В данной главе дается описание работы оптических бистабильных устройств и логических вентилях. Подводятся итоги последних достижений, обсуждаются три вида перспективных устройств, использующих GaAs. Дается сравнение чисто оптических устройств с одной из гибридных систем, а именно с

устройством на собственном электрооптическом эффекте (СЭОУ) [2]. Обсуждаются предельные возможности этих устройств с точки зрения потребляемой мощности, быстродействия, плотности упаковки и статистических свойств света. Выделены такие существующие в настоящее время проблемы управления передачей информации светом, как дифракция, рассеяние, каскадирование, тепловыделение, а также их влияние на характеристики нелинейных устройств. Вскользь упомянуты схемы архитектуры компьютеров, например трехуровневые системы. В заключение обсуждаются перспективы развития данной области.

2.2. Принципы работы бистабильных устройств

Оптическая бистабильность требует применения нелинейных материалов и оптической обратной связи. В тех устройствах модуляции света, где используют изменение показателя преломления, применяют нелинейные оптические среды, имеющие показатель преломления, зависящий от интенсивности света. Обратная связь является или внешней (макроскопической), где нелинейная среда размещается внутри интерферометра (резонатора) Фабри—Перо¹⁾, или внутренней (микроскопической), где оптически индуцируемые изменения в нелинейной среде непосредственно влияют на взаимодействие среды с падающим пучком света. Большая часть работ по оптическим переключениям до сих пор выполнялась на устройствах, где для получения бистабильности использовались изменения рефрактивных свойств материалов, а обратная связь осуществлялась с помощью внешнего резонатора Фабри—Перо. Примером реализации внутренней обратной связи является случай бистабильности, обусловленной возрастанием коэффициента поглощения. В устройствах типа СЭОУ (обсуждаемых ниже), хотя и используют рост коэффициента поглощения, но для воздействия на поглощение применяют извне подаваемое электрическое поле, так что здесь обратная связь является внешней. До сих пор все из наиболее перспективных устройств основывались на внешней обратной связи.

Резонатор Фабри—Перо состоит из двух частично отражающих зеркал; он пропускает свет только тогда, когда оптическая длина пути (физическое расстояние между зеркалами, умноженное на показатель преломления среды) равняется целому числу, умноженному на половину длины волны света. Следовательно, выходной сигнал резонатора Фабри—Перо

¹⁾ Называемого также эталоном (см. Гиббс Х. Оптическая бистабильность. Управление светом с помощью света: Пер. с англ. — М.: Мир, 1988), — *Прим. перев.*

представляет собой серию острых максимумов пропускания, причем для каждого из пиков выполняется условие резонанса. Для работы в бистабильном режиме длину волны проходящего светового пучка вначале каким-либо способом отстраивают относительно максимума пропускания резонатора Фабри—Перо; коэффициент пропускания при этом является низким и устройство выключено. При увеличении интенсивности показатель преломления материала, заполняющего резонатор, изменяется, изменяя тем самым оптическую длину пути, и сдвигает пик пропускания, приводя к более точному совпадению длины волны максимума пропускания и длины волны лазера. Это в свою очередь увеличивает интенсивность света в резонаторе, которая далее изменяет показатель преломления, и т. д. Эта положительная обратная связь сохраняется до критической величины интенсивности, называемой интенсивностью включения. Затем эффект пропадает и устройство быстро переходит во включенное состояние. Дальнейшее увеличение интенсивности света не приводит к значительному изменению коэффициента пропускания. Если же интенсивность света снижать ниже интенсивности включения, коэффициент пропускания не умень-

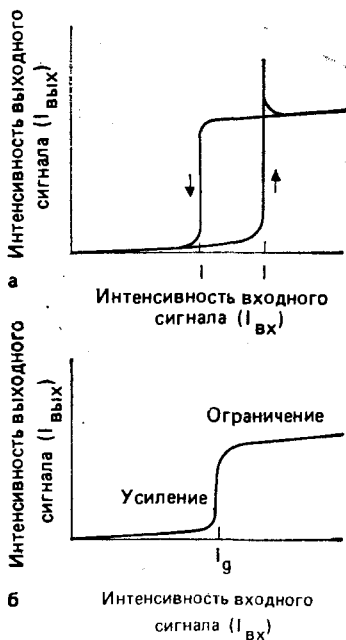


Рис. 2.1. Зависимости выходного сигнала от входного сигнала для нелинейного резонатора Фабри—Перо, демонстрирующего: а) оптическую бистабильность; б) пороговую зависимость.

шится, а останется в «верхнем» состоянии. Это происходит потому, что вследствие свойства накопления поля в резонаторе Фабри — Перо интенсивность света внутри резонатора остается высокой. Устройство выключается, когда интенсивность уменьшается до значения, при котором интенсивность внутри резонатора становится ниже, чем требуется для поддержания устройства во включенном состоянии. На рис. 2.1, а показана характеристика пропускания света бистабильного резонатора Фабри — Перо. Гистерезисный характер кривой обусловлен процессом, описанным выше. При определенных условиях работы эту гистерезисную область можно исключить, и характеристика пропускания станет близкой к приведенной на рис. 2.1, б. Как будет показано дальше, для работы оптических вентилях устройств не требуется бистабильности, и достаточной является нелинейная характеристика, приведенная на рис. 2.1, б. То же самое переключающее поведение может быть получено при использовании двух световых пучков (рис. 2.2). Интенсивный лазерный пучок, называемый несущим пучком, изначально отстроен по длине волны относительно максимума коэффициента пропускания, что определяет выключенное со-

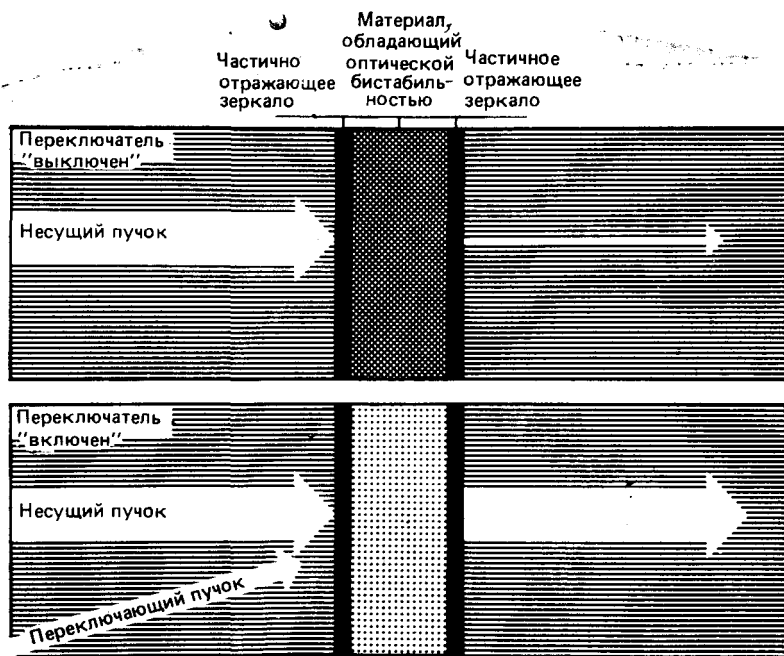


Рис. 2.2. Двухлучевое переключение в нелинейном бистабильном резонаторе Фабри — Перо.

стояние. Когда же приложен переключающий пучок с низкой интенсивностью, устройство включается и несущий пучок проходит через устройство. Переключающий пучок в этом случае содержит вводимую в устройство информацию.

2.3. Требования к характеристикам нелинейных устройств и материалам

Оптические бистабильные устройства и логические элементы, чтобы не возникало проблем из-за выделяемого тепла, должны при выполнении большого числа параллельных операций работать с малыми затратами мощности. Расчеты, проведенные с учетом статистических свойств света, показывают, что необходимо по крайней мере примерно 300 фотонов на бит и тогда частота появления ошибок будет менее 10^{-12} . При этом, например, затраты энергии для устройства, работающего в видимом диапазоне длин волн, составят приблизительно 0,1 фДж. А при условии, что мы хотим достичь скоростей переключения порядка 1 пс, величина потребляемой мощности должна составлять 0,1 мВт (если, конечно, такие энергии и скорости достижимы). Устройства должны иметь малые размеры (несколько квадратных мкм), работать при удобных значениях температуры (например, комнатной) и длины волны.

Указанные выше желательные параметры нелинейных устройств в свою очередь накладывают определенные требования на материалы. Так, для переключения при малых мощностях необходимо наличие у материалов сильных нелинейных свойств (больших n_2). Быстрый отклик и малые времена релаксации нелинейной среды позволили бы достичь коротких циклов переключения. Нелинейность должна существовать при комнатной температуре. Полупроводники и до некоторой степени органические и фоторефрактивные материалы удовлетворяют большинству этих требований. Однако материал, который удовлетворял бы одновременно всем этим требованиям, еще не найден. Электронные и оптические свойства полупроводников, используемых в нелинейных устройствах, можно изменить, если воспользоваться особыми свойствами электронов проводимости в полупроводниках, облучаемых квантами света с энергиями, близкими энергии ширины запрещенной зоны. В этом случае многие полупроводники, а в частности GaAs, пригодны для создания электронных, оптических или оптоэлектронных устройств. Полупроводники также вполне удовлетворяют требованиям, предъявляемым к нелинейным резонаторам Фабри—Перо. Они привлекают тем, что дают возможность получить соответствующий коэффициент поглощения (будучи умноженным на длину он составляет $\alpha L \simeq 1$) на очень коротких расстояниях (~ 1 мкм), что позволяет создавать

устройства очень маленьких размеров. Малая длина нелинейного резонатора Фабри — Перо обеспечивает малое время полного обхода резонатора светом, так что предельные времена переключения, обусловленные временем нарастания импульса в резонаторе, составляют обычно субпикосекунды ($\sim 10^{-11}$ с). Малая длина резонатора также обеспечивает более высокую степень фокусировки света без существенного увеличения потерь на расходимость, что позволяет снизить мощность и энергию переключения. Кроме того, чем лучше сфокусировано лазерное излучение, тем меньше может быть сделан размер переключающего устройства, что справедливо в том случае, если, например, можно исключить диффузию носителей за счет подбора размеров ячейки (см. ниже).

Существование резонансных эффектов в полупроводниках, характеризуемых большими оптическими нелинейностями, позволяет получить малые мощности переключения. Например, в квантоворазмерных структурах¹⁾ на GaAs—AlGaAs нелинейная добавка к показателю преломления n_2 , связанная с резонансом свободных экситонов при комнатной температуре, составляет $\sim 0,1$ см²/кВт [4]. Полупроводники имеют довольно малые времена релаксации, соответствующие нано- и субнаносекундным временным интервалам, что обуславливает малые времена переключения. Увеличенная поверхностная рекомбинация уменьшила время жизни носителей и время восстановления вентиля HE, построенного на резонаторе Фабри — Перо, заполненном GaAs, до величин менее 200 пс [5]. Облучение сверхрешеток пучками протонов [6] также позволяло получить близкие времена релаксации. Некоторые полупроводниковые устройства, такие как квантоворазмерные структуры на GaAs и AlGaAs, сохраняют свои нелинейные свойства при повышенных температурах [7] и позволяют достичь переключения при комнатной температуре [8, 9].

2.4. Причины возникновения нелинейных оптических эффектов в полупроводниках

Механизм нелинейности в полупроводниках связан с характером оптического возбуждения в материале. При поглощении квантов с энергией, меньшей ширины запрещенной зоны, в полупроводнике образуются экситоны [10]. Свободный экситон представляет собой связанное состояние электрона и дырки (связь осуществляется посредством кулоновского взаимодейст-

¹⁾ Дословный перевод английского термина MQW (multiple quantum well structures) — структура с большим числом квантовых ям. В отечественной литературе используются одновременно термины *сверхрешетка* и *квантово-размерная структура*. — *Прим. ред.*

вия). Такой экситон возникает в полупроводнике при поглощении кванта света с энергией, близкой ширине запрещенной зоны. Некоторые экситоны могут быть локализованы на примесях, обычно присутствующих в полупроводниках (донорных или акцепторных), образуя таким образом связанный экситон [11]. Биэкситон — это связанное состояние двух экситонов, несколько напоминающее молекулу водорода, являющуюся связанным состоянием двух атомов водорода [10]. Экситонная нелинейность возникает вследствие насыщения экситонного поглощения и соответствующего изменения показателя преломления. Это «насыщение» обусловлено тем, что присутствие либо уже возникших экситонов, либо свободных носителей затрудняет образование экситона. Следовательно, интенсивность света, необходимая для выполнения переключений в экситонных системах, тесно связана с величиной интенсивности света, необходимой для насыщения резонанса экситонного поглощения. Интенсивность насыщения составляет меньшую величину для экситона большего радиуса или меньшей энергии связи. Более слабое электрон-фононное взаимодействие уменьшает эффект уширения линии экситонного поглощения и является причиной того, что эффекты, связанные с экситонами, наблюдаются при больших температурах. Следовательно, умеренные энергии связи экситонов вместе с минимальным электрон-фононным взаимодействием являются идеальным условием для реализации переключающих операций, характеризующихся малыми затратами мощности при комнатной температуре. Другой нелинейный механизм — это эффект «заполнения зоны», при котором заполнены состояния с энергиями, большими ширины запрещенной зоны, что препятствует поглощению на этих энергиях. Соответственно спектральная зависимость поглощения оказывается сдвинутой, и при энергиях, как раз меньших ширины запрещенной зоны, могут возникать большие изменения показателя преломления. В полупроводниках часто сосуществуют нелинейности двоякой природы — связанные как с экситонами, так и с заполнением зоны. Хотя сказанное здесь является крайне упрощенным описанием оптических нелинейностей в полупроводниках, оно содержит существенные элементы для понимания основ этих явлений.

2.5. Современное состояние работ по оптическим бистабильным устройствам

Оптическая бистабильность наблюдалась во многих полупроводниках, таких как GaAs [8, 9] (бистабильность на свободных экситонах), GaAs—AlGaAs (квантоворазмерные структуры) [2, 8—10], CuCl [12, 13] (биэкситонная бистабильность), CdS [11] (бистабильность на связанных экситонах), InSb [14,

15] (бистабильность на заполнении зоны), InAs [16] (бистабильность на заполнении зоны), CdHgTe [17] (бистабильность на заполнении зоны), ZnS и ZnSe [18—20] (тепловая бистабильность). В следующем разделе сделан обзор рабочих характеристик бистабильных устройств и логических вентилях на квантоворазмерных структурах из GaAs и AlGaAs. С характеристиками других полупроводников можно ознакомиться в монографии [1] или обзорах [10, 21]. Оптическая бистабильность при комнатной температуре наблюдалась как в объемных структурах на GaAs, так и на GaAs-квантоворазмерных структурах (рис. 2.3) [8]. В работе [9] сообщается о снижении потерь мощности на переключение до 5 мВт при комнатной температуре. В качестве источника света для переключений при комнатных температурах в структурах с резонатором Фабри—Перо с промежуточным слоем квантоворазмерной структуры с толщиной слоев GaAs в 5,3 нм, AlGaAs в 5,6 нм использовался полупроводниковый лазер (рис. 2.4) [22].

2.6. Оптические логические элементы Фабри—Перо (ОЛЭФП)

Используемые в экспериментах по бистабильности и описанные выше нелинейные элементы Фабри—Перо способны выполнять большое число различных логических операций [23, 24]. Процедура получения логических операций достаточно наглядно проявляется при рассмотрении эксперимента по накачке и зондированию устройства. Сигнал накачки при сравнительно низких интенсивностях сильно взаимодействует с нелинейной средой, в то время как высокоинтенсивный зондирующий сигнал, сравнительно слабо взаимодействующий со средой, настроен на область максимума пропускания. Сдвиг этого пика пропускания, вызванный накачкой входным пучком, изменяет пропускание зондирующего луча и в зависимости от начальной степени отстройки зондирующего пучка относительно максимума пропускания (рис. 2.5) может приводить к выполнению таких логических функций, как ИЛИ-НЕ, И-НЕ, ИСКЛЮЧАЮЩЕЕ ИЛИ, ИЛИ, И. Например, когда зондирующий луч изначально (т. е. в отсутствие входного сигнала) смещен вправо от максимума пропускания, реализуется функция ИЛИ-НЕ, потому что уровень входного сигнала «1» или «2» смещает максимум в сторону от длины волны зондирования. Коэффициент пропускания зондирующего пучка определяет выходной сигнал, поэтому он будет низким в обоих случаях.

Логические вентили на основе ОЛЭФП наиболее эффективно реализуются тогда, когда и входной, и зондирующий сигналы представляют собой импульсы с длительностью много меньшей, чем время релаксации среды. Это уменьшает затра-

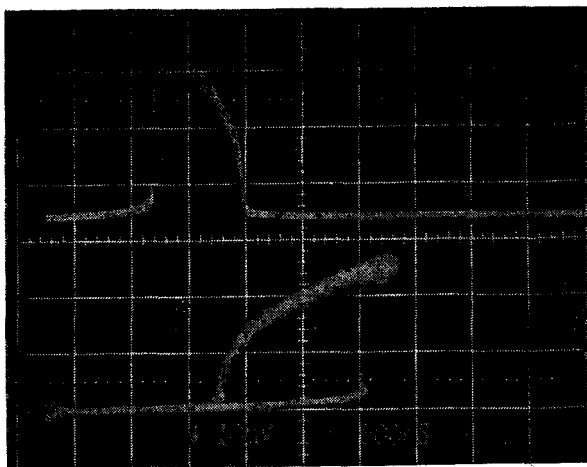


Рис. 2.3. Оптическая бистабильность в квантоворазмерных структурах на GaAs [8].

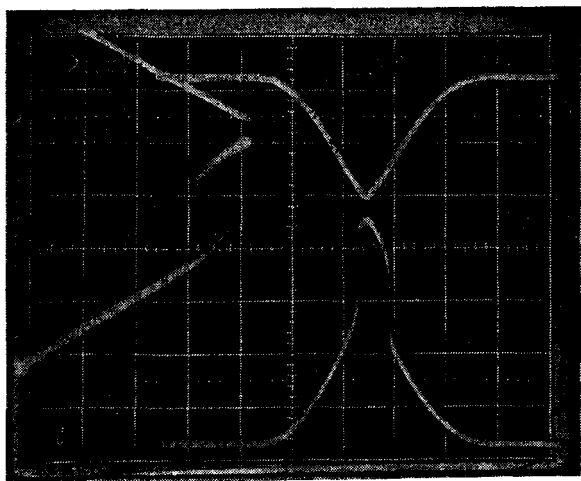


Рис. 2.4. Бистабильность в резонаторе Фабри — Перо, заполненном квантоворазмерной структурой на GaAs, на длине волны излучения полупроводникового инжекционного лазера 830 нм. Верхняя правая развертка: инвертированная зависимость входного сигнала от времени (максимум около 6 мВт). Верхняя левая развертка: зависимость входного от входного сигнала. Нижняя левая развертка: зависимость выходного сигнала от входного. Нижняя правая развертка: зависимость выходного сигнала от времени. Временной масштаб: 500 нс/дел [21].

ты энергии в пересчете на одну логическую операцию, поскольку атомы, которые должны быть возбуждены, чтобы получить достаточно сильный нелинейный эффект, нуждаются лишь в однократном возбуждении. Бистабильный режим работы устройства или режим работы с выдачей непрерывного сигнала, характеризуемого, кроме того, усилением сигнала, будет поддерживаться по крайней мере на протяжении среднего времени релаксации среды, таким образом за время выполнения логической операции атомы релаксируют и повторно возбуждаются. Время полного цикла выполнения операции в импульсном режиме также сведено до минимума, потому что в этом случае отсутствует постоянная засветка на входе, и релаксация происходит в темноте, с минимальным временем релаксации.

Для того чтобы получить большой коэффициент усиления за времена, много меньшие среднего времени релаксации, или достичь чувствительности нелинейной среды к входному пучку, а не к зондирующему сигналу, устройство должно каким-то образом различать входной и зондирующий лучи. Для этого могут быть использованы такие параметры, как длина волны, поляризация, а также различия в геометрии прохождения

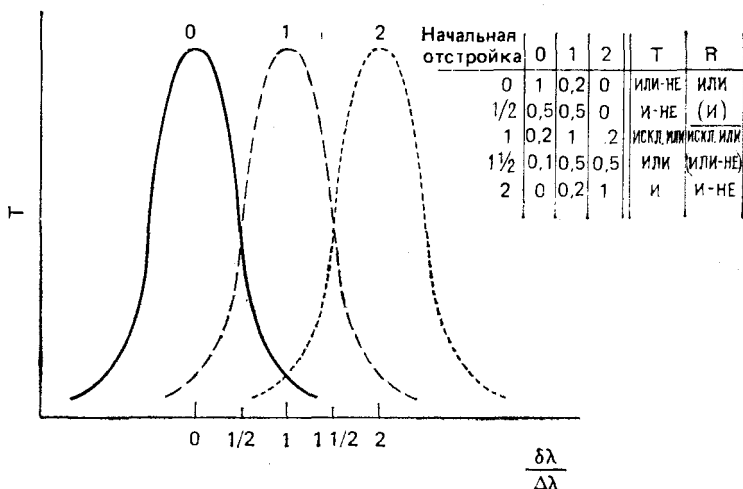


Рис. 2.5. Положение пика пропускания для величин входного сигнала в 0, 1 или 2 (показано сверху над пиками). Начальное значение отстройки резонатора, отложенное по горизонтальной оси, выражено в единицах ширины пика пропускания на полувысоте. Указанные в таблице логические функции получены при пяти значениях отстройки. Дробные значения в столбцах таблицы под значениями входного сигнала (0, 1, 2) представляют собой приближенные значения пропускания зондирующего излучения, когда каждый входной сигнал сдвигает пик пропускания на одну ширину. Показаны значения функций вентиля, полученные при пропускании и отражении зондирующего луча [23].

света или во времена прохода света в материале. Селективность по длине волны привлекает тем, что она часто реализуется очень простыми средствами. При этом для дальнейшей оптимизации как устройства, так и всей системы остается возможность использовать поляризацию, длину волны и т. д. Входной и выходной сигналы в данном случае соответствуют разным длинам волн, поэтому при необходимости реализации в схеме петли обратной связи на материал накладываются серьезные дополнительные ограничения. (Данный вывод касается любого устройства, которое преобразует выходной сигнал в длину волны входного сигнала.) При этом во втором вентиле предпочтительно использовать нелинейный материал со свойствами, обратными к свойствам материала в «первом» устройстве. Другими словами, желательной является чувствительность к длине волны первого зондирующего луча, а не к длине волны первого входного луча. В этом случае действительно возможно построение замкнутой схемы, содержащей два взаимно обратных логических вентиля. Для задач, требующих нескольких шагов обработки данных, можно построить каскад резонаторов Фабри—Перо, заполненных полупроводниковыми материалами, с последовательно увеличивающимися длинами волн.

Режим работы с двумя длинами волн позволяет провести эффективную оптимизацию схемы с помощью двух зеркал, имеющих большой коэффициент отражения на длине волны зондирующего луча (чтобы получить хорошую настройку резонатора), но пропускающих входной луч [24]. Такая конструкция при заполнении резонатора слоем GaAs или квантово-размерными структурами на GaAs позволила при выполнении логических операций получить контраст 5:1 при энергии входного сигнала лишь 3 пДж [24]. Для чисто оптического логического устройства это представляет собой минимальное из значений энергии переключения, о которых когда-либо сообщалось в печати; однако сюда не включена «неактивная» энергия зондирующего луча, величина которой определяется коэффициентом пропускания устройства и требованиями к усилению выходного сигнала. Энергия зондирующего луча примерно в 10 раз больше энергии входного луча, но поглощение последнего должно быть мало. На рис. 2.6 с целью демонстрации релаксационных характеристик показаны функции отклика устройства, на вход которого подан сигнал в 8 пДж (импульсы генерируются в режиме с синхронизацией мод); представленные зависимости соответствуют случаю непрерывного облучения устройства зондирующим лучом. Если импульс зондирующего излучения подается сразу же за импульсом входного сигнала, то на него не будут оказывать влияние изменения максимума пропускания, происходящие в процессе ре-

лаксации при переходе через значение, равное длине волны зондирующего луча. Например, в элементе ИСКЛЮЧАЮЩЕЕ ИЛИ при подаче двух импульсов входного сигнала зондирующий импульс отследил бы только состояние, возникающее после подачи входного импульса, соответствующее минимуму пропускания, а не состояние среды, соответствующее максимуму пропускания. В работе [25] описаны эксперименты, в которых время отклика GaAs-вентилей ИЛИ-НЕ определялось

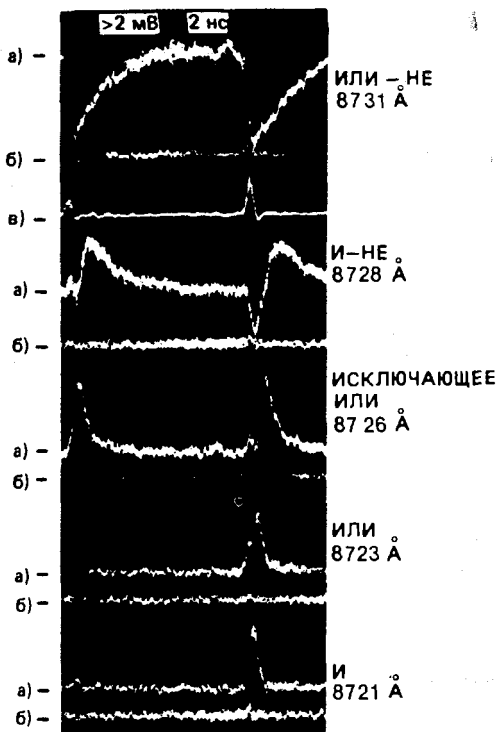


Рис. 2.6. Экспериментальные результаты, полученные с помощью оптических логических элементов (ОЛЭФП):

а) зависимость пропускания зондирующего пучка от времени при наличии входных сигналов. Отклик слева обусловлен входным импульсным сигналом в 8 пДж (входной сигнал —1); отклик с правой стороны обусловлен входным сигналом в 16 пДж (входной сигнал —2);

б) те же зависимости, что и для а), при перекрытом зондирующем пучке (нулевой уровень сигнала);

в) отраженные сигналы от входных импульсов при перекрытом зондирующем луче. Отраженные сигналы измеряются тем же самым фотодиодом, при введенной в систему четвертьволновой пластинке. Данная методика обеспечивает точную синхронизацию разверток а) и в). Входные импульсы были одинаково для всех вентилях и показаны только для случая вентиля ИЛИ-НЕ [24].

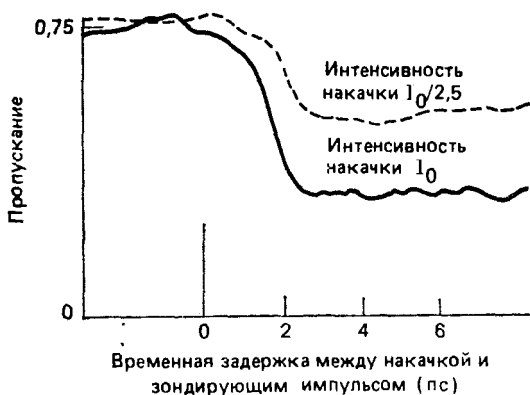


Рис. 2.7. Временной отклик в ~ 1 пс для вентиля ИСКЛЮЧАЮЩЕЕ ИЛИ, использующего квантоворазмерную структуру на GaAs [25].

с помощью фемтосекундной лазерной системы. Временная характеристика, соответствующая времени переключения бистабильного устройства, может быть получена путем измерения того, насколько быстро сдвигается пик пропускания резонатора Фабри—Перо в ответ на воздействие входного пучка. Полученное для работающего при комнатной температуре устройства время отклика ~ 1 пс является минимальным для оптических логических устройств с малым потреблением мощности (рис. 2.7). Время возврата вентиля в исходное состояние, которое соответствует времени выключения бистабильного уст-

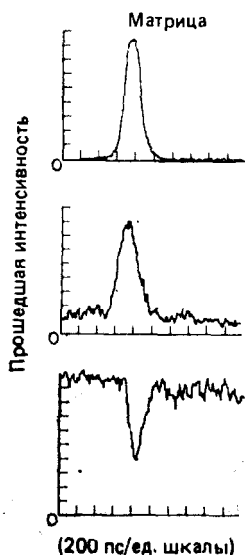


Рис. 2.8. Время восстановления вентиля ИСКЛЮЧАЮЩЕЕ ИЛИ, использующего квантоворазмерную структуру на GaAs без AlGaAs «окон». Устройство демонстрирует времена восстановления < 200 пс (ограничения связаны с измерительной системой). Увеличенная поверхностная рекомбинация привела к меньшим временам жизни носителей и, следовательно, меньшим временам восстановления элемента [5].

ройства (без применения таких специальных методик уменьшения времени выключения, как увеличение поверхностной рекомбинации или протонная бомбардировка), составляло несколько наносекунд вследствие больших времен жизни носителей (рис. 2.6). Однако увеличение поверхностной рекомбинации, получаемое при удалении верхнего «окна» из AlGaAs, позволило достичь меньшего времени жизни носителей и времени восстановления для устройства на GaAs менее 200 пс (рис. 2.8) [5]. Реально это устройство представляет собой двумерную матрицу 100×100 ячеек, при этом каждая ячейка имеет размеры 9×9 мкм [26]. На рис. 2.9 показан фрагмент матрицы. Резонатор Фабри—Перо толщиной 0,5 мкм, заполненный GaAs и не имеющий «окон» из AlGaAs, демонстрирует фактически полное восстановление пропускания за 150 пс [27] (рис. 2.10) при энергии входного импульса в 16 пДж. Данные цифры указывают на то, что в потенциале возможен режим работы с частотой переключения более 5 ГГц. Все эксперименты с ОЛЭФП до настоящего времени проводились при комнатной температуре и в них преобразовывали коротковолновое (с энергией, превышающей ширину запрещенной зоны) входное излучение в более длинноволновое (с энергией, меньшей ширины запрещенной зоны). Процессы преобразования меньшей частоты в большую частоту, по-видимому, возможны в охлаждаемых элементах на GaAs, а также при заполнении ре-

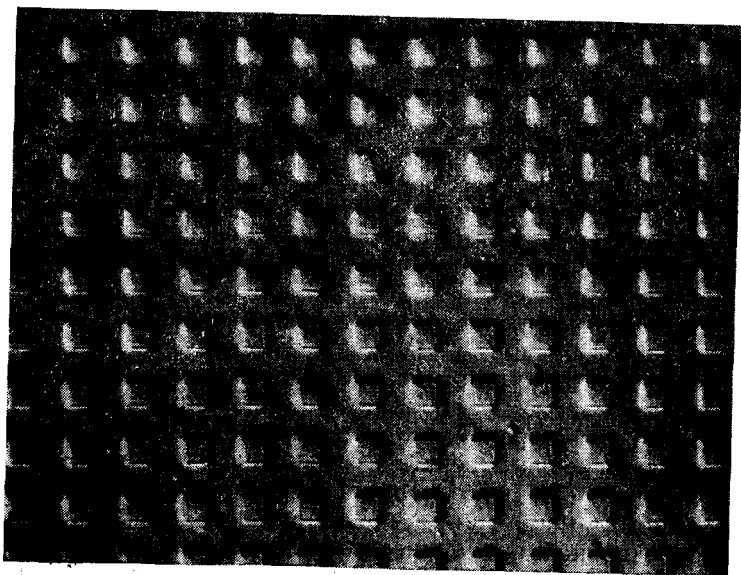


Рис. 2.9. Фрагмент матрицы 100×100 ячеек на GaAs, изготовленной плазменным травлением. Каждая ячейка имеет размер 9×9 мкм [2].

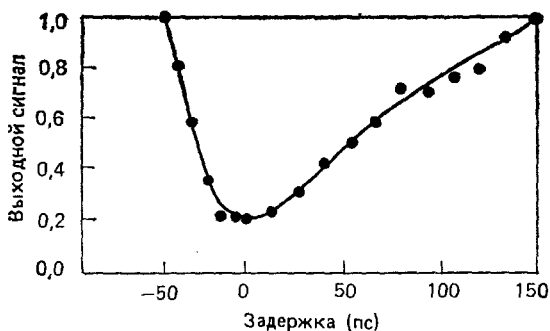


Рис. 2.10. Пропускание вентиляющего элемента ИЛИ-НЕ в зависимости от времени задержки зондирующего сигнала. Из зависимости следует, что полное время восстановления элемента около 150 пс.

зонатора CdS и охлаждении элемента до температуры жидкого гелия [11], т. е. там, где удается получить обособленный пик поглощения. Такие процессы также должны быть возможными при комнатной температуре в случае использования переходов между состояниями в квантовых ямах в квантоворазмерных структурах на GaAs [28]. Дальнейшая работа по улучшению характеристик материалов и использованию новых методов может существенно упростить задачу каскадирования резонаторов Фабри—Перо, работающих на двух частотах, что в настоящее время представляет собой весьма сложную задачу.

Важная особенность, отличающая ОЛЭФП от бистабильных устройств, состоит в том, что они являются (используя обиходный термин из электротехники) *трехполюсниками*, т. е. ОЛЭФП «изолирован» от выхода, поскольку любая обратная связь не оказывает на него значительного влияния. Это свойство может иметь достаточно большое значение в реальных системах, где неизбежно присутствует определенное влияние обратной связи. Более того, ОЛЭФП могут быть просто блокированы от света даже на длине волны входного сигнала, если он почему-либо распространяется в устройстве в обратном направлении, т. е. от выхода ко входу устройства, что позволяет снизить требования, предъявляемые к величине входного сигнала. Схема, названная «диод данных» (*data-diode*), имеет входное пропускающее зеркало только на одной стороне, в то время как другое зеркало полностью отражает входной сигнал. Таким образом, падающий по заданному направлению луч совершает двойной проход через нелинейную среду, в то время как аналогичный сигнал, поступающий в схему по какой-либо причине с противоположной стороны, отражается, не попадая во внутрь. Такое недавно опробованное устройство при

правильной настройке продемонстрировало контраст 5:1, но потребовалось удвоить энергию входного сигнала, чтобы получить хотя бы заметный отклик при перевороте устройства другой стороной.

2.7. Устройства на собственном электрооптическом эффекте (СЭОУ)

Устройства на собственном электрооптическом эффекте [2] (СЭОУ) являются еще одним видом бистабильных устройств, работа которых отличается от нелинейных резонаторов Фабри—Перо в двух важных аспектах. Во-первых, прохождение света в материале увеличивает коэффициент поглощения света этого же материала на рабочей длине волны. Именно из-за возрастающего поглощения света под действием света этот класс бистабильности обычно называется оптической бистабильностью [29]. В этом процессе благодаря эффекту поглощения света увеличивается удельный коэффициент поглощения света, что приводит к возникновению положительной обратной связи, которая переводит режим прохождения света из состояния с высоким пропусканием в состояние с низким пропусканием, т. е. противоположно тому, как это происходит в устройствах на основе резонатора Фабри—Перо. Указанные устройства не требуют применения резонатора и не имеют ограничений быстродействия, связанных с временем нарастания импульса в резонаторе. В них используются особенности резонансного поведения коэффициента поглощения, которые ограничивают ширину полосы частот (хотя и не жестко). Второе отличие СЭОУ состоит в том, что удельный коэффициент поглощения изменяется в соответствии с изменением приложенного электрического поля, вызывающего генерацию свободных носителей. Как было показано ранее, квантоворазмерные структуры на основе GaAs—AlGaAs очень чувствительны к электрическим полям, приложенным перпендикулярно к слоям, образующим квантовые ямы. Сдвиг края поглощения, вызванный небольшими изменениями электрического поля, в таких квантоворазмерных структурах называется эффектом Штарка для носителей заряда, удерживаемых в квантовой яме [30]. Такой эффект оказывает очень сильное воздействие, в результате которого СЭОУ переключаются при мощности оптического излучения лишь 4 фДж/мкм² по сравнению с достигнутыми до сих пор ~ 30 фДж/мкм² (только для входного сигнала) для ОЛЭФП. Однако дополнительно СЭОУ поглощают электрическую мощность ~ 14 фДж/мкм². (Кроме того, отметим, что представляется маловероятным получить СЭОУ таких же малых размеров, как устройства на резонаторах Фабри—Перо.) Таким образом, полная поглощаемая энергия

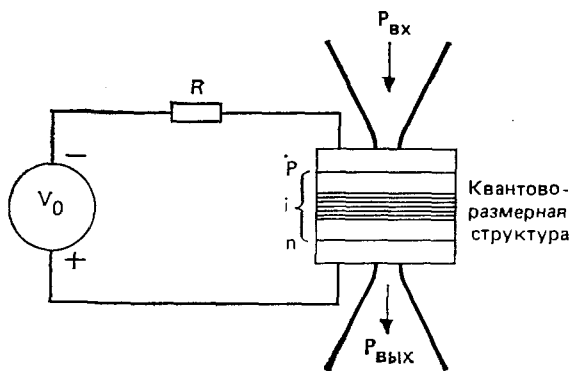


Рис. 2.11. Общая схема работы СЭОУ [2].

СЭОУ, по-видимому, является более высокой. Для того чтобы приложить электрическое поле в поперечном направлении к кванторазмерным структурам, имеющим пренебрежимо низкие токи (в отсутствие свободных носителей), используют схему с указанной на ней полярностью включения элемента (рис. 2.11). На рис. 2.12 показаны характеристики СЭОУ, имеющего диаметр ~ 100 мкм и управляемого при комнатной температуре энергией света ~ 100 пДж [2]. Характеристики переключения в обоих направлениях не идентичны. Величина контраста ограничена исключительно сильным поглощением в сильно энергетической области резонанса поглощения (тот же механизм поглощения вызывает трудности с каскадированием ОЛЭФП на основе GaAs при комнатной температуре). Хотя быстродействие СЭОУ не ограничено потерями времени на нарастание поля в резонаторе, оно, однако, ограничено временной RC-постоянной, определяемой управляющей электронной схемой. В настоящее время самый короткий экспериментально полученный временной цикл составляет около 30 нс, но в устройствах в интегральном исполнении возможны субнаносекундные времена.

Несмотря на ограничения, накладываемые на размер, быстродействие и контраст, прошедшие испытания несколько типов СЭОУ проявили себя исключительно хорошо. Как и в случае с устройствами на основе резонаторов Фабри—Перо, дальнейшие разработки в области материалов могут привести к значительному улучшению характеристик устройств данного типа. Характерной чертой, которая отличает СЭОУ от чисто оптических устройств, является то, что они могут быть сопряжены с электроникой сразу двумя путями; т. е. ими можно управлять чисто электронными способами, используя в качестве модуляторов света, или они могут быть включены в схему в качестве

фотодетекторов. Это их качество обоснованно делает вероятным то, что СЭОУ или какие-то ответвления от этой основной концепции будут играть важную роль в оптических вычислительных системах.

2.8. Факторы, ограничивающие предельные величины коэффициентов усиления, разветвления по выходу, объединения по входу

Оптические бистабильные устройства (ОБУ) (резонаторы Фабри—Перо, СЭОУ и др.) имеют очень большие, стремящиеся к бесконечности значения дифференциальных коэффициентов усиления при переключающей интенсивности $I_{вкл}$. В литературе, например [31], встречались сообщения об экспериментально полученных коэффициентах усиления, превышающих 10^4 . Здесь для простоты и наглядности под коэффициентом усиления, называемым далее просто усилением, подразумевается крутизна зависимости выходного сигнала от входного сигнала.

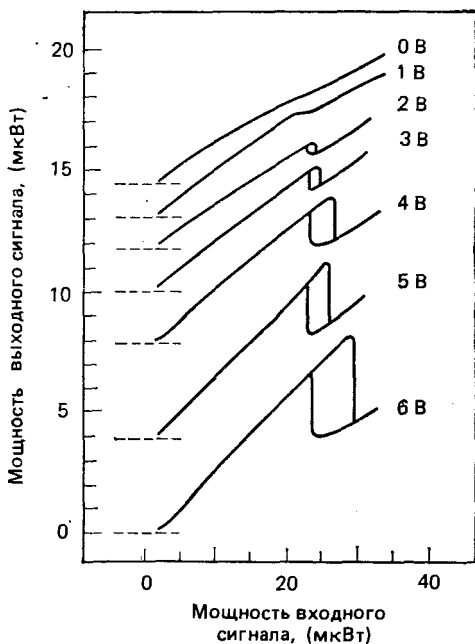


Рис. 2.12. Интенсивность на выходе СЭОУ в зависимости от входной интенсивности для $\lambda=853$ нм. Диаметр СЭОУ — 100 мкм, напряжения питания обратной полярности — от 0 до 6 В, используется режим питания постоянного тока при 9,3 мкА. Для наглядности нулевые значения последовательно смещены относительно друг друга.

Однако имеется ряд соображений, ограничивающих получаемое на практике значение коэффициента усиления ОБУ. Так, если входной сигнал будет незначительно превосходить $I_{вкл}$, тогда время переключения будет существенно больше времени релаксации среды. Это явление, называемое критическим замедлением [32], возникает вследствие установления динамического равновесия между процессами возбуждения и релаксации атомов. Таким образом, чтобы использовать быстродействие, присущее ОБУ, и получить большую скорость переключения, необходимо, чтобы входной сигнал имел заметно большую интенсивность по сравнению с $I_{вкл}$. Далее, такие факторы, как нестабильность лазерного излучения $I_{вкл}$ (например, вследствие изменений температуры и неоднородности характеристик материалов), разброс значений коэффициента пропускания системы и возможного влияния старения прибора, диктуют необходимость увеличения входных сигналов по отношению к подпороговому значению интенсивности. Таким образом, получаемое на практике полезное усиление определяется скорее компонентами системы и соображениями надежности, чем параметрами самого нелинейного устройства.

Значение коэффициента усиления ОЛЭФП зависит преимущественно от параметров устройства. В данном случае коэффициент усиления определяется как отношение приращения энергии выходного сигнала к приращению энергии входного пучка света. Единственным фактором, ограничивающим время одиночного цикла переключения, является время нарастания поля в резонаторе. Обсуждавшиеся ранее в отношении ОЛЭФП факторы влияют на однородность выходного сигнала, но прямую усиление не ограничивают. Так, если бы зондирующий пучок действительно не влиял на пропускание ОЛЭФП, то можно было бы получить произвольно большие значения коэффициента усиления, так как энергия зондирующего пучка при этом могла быть произвольно большой. Этого нет на практике, и именно взаимодействие зондирующего пучка с нелинейной средой ограничивает получаемое усиление. Например, в логическом венти́ле ИЛИ-НЕ в отсутствие входного сигнала, если энергия зондирующего пучка крайне велика, именно это взаимодействие будет определять процесс возбуждения среды и соответственно сдвиг пика пропускания устройства относительно длины волны света. Следовательно, энергия выходного пучка не может быть сделана произвольно большой. Это ограничение усиления аналогично ограничению мощности в рефрактивных ОБУ на основе резонатора Фабри—Перо и может быть использовано для уменьшения шума в системе. В этом случае максимальное усиление приблизительно определяется коэффициентом пропускания зеркала, разделенным на удвоенную длину поглощения света (все параметры относятся

к длине волны зондирующего пучка) [33]. Усиление и коэффициент разветвления по выходу связаны между собой очень просто.

Усиление должно быть по меньшей мере таким же, как коэффициент разветвления по выходу, деленный на коэффициент пропускания системы от одного до следующего устройства.

Способы реализации разветвления по выходу и объединения по входу тесно связаны с архитектурой системы. Однако здесь они не рассматриваются. Отметим лишь, что наиболее желательным является значение коэффициента объединения по входу, равное 2. Данное утверждение определяется следующими причинами: (1) Чем меньше входов у вентиля, тем он проще и надежнее. Именно по этой причине почти все существующие логические устройства имеют два входа. (2) Коэффициент объединения по входу, равный 2, позволяет избежать ряд проблем, связанных с интерференцией входных сигналов, имеющих разную поляризацию. (3) В планарных матричных системах изображений с помощью ортогональных поляризаций несложно получить коэффициенты разветвления и объединения, равные 2. Для этого следует разместить призму Волластона в коллимирующем устройстве системы.

2.9. Концепции архитектуры

Планарно-матричные архитектуры обработки изображений, т. е. переноса одной матрицы изображения на другую преобразующую матрицу, с успехом реализуются на основе устройств, рассмотренных выше. Планарно-матричная архитектура обеспечивает полную параллельность обработки массива (основное соображение в пользу оптических вычислений) и использует хорошо развитые методы классической оптики [34]. При этом линза становится межэлементным соединением сразу для 10^8 элементов/см² с субфемтосекундной временной однородностью. Обычно упорядоченность таких соединений рассматривают в качестве механизма ограничения степени универсальности такого оптического компьютера. Однако даже в электронных цепях по мере увеличения их быстродействия становится все более необходимым сделать одинаковыми длины проводов, соединяющих элементы. Данное требование совместно с недопустимостью пересечения проводов заставляет использовать в электронике все более и более упорядоченные соединения. Оказывается, что необходимость в упорядоченности соединений обуславливается в основном стремлением добиться высокой скорости и простоты изготовления, а не типом используемых логических элементов. За последнее время сделаны значительные успехи в области разработки архитектур обработки изображений. Символьные подстановки [35] стали од-

новременно простым и мощным методом оптической обработки информации, позволяющим распознавать много фрагментов в большой матрице данных и заменять их другими фрагментами. Наряду с этим при использовании упорядоченных соединений с матрицами логических вентилей ИЛИ-НЕ или И возможна также реализация арифметической [35] и других видов обработки. В работе [35] также описан простейший компьютер, основанный на символьных подстановках.

2.10. Предельные параметры и вычислительные возможности ОБУ

При выполнении заданных логических операций оптическим устройством оптические характеристики нелинейного материала должны измениться на определенную величину. Это находит свое отражение в том, что уровень возбуждения в материале, например плотности экситонов или свободных носителей, достигает определенного уровня. Отсюда следует, что ключевым подходом к минимизации затрат энергии на переключение является сведение объема материала к минимуму. Данная стратегия, кроме того, приводит к повышению быстродействия и степени интеграции. Однако здесь будут возникать трудности чисто технологического характера, если качество материала не отвечает предъявляемым требованиям. Тем не менее в устройствах на основе резонаторов Фабри—Перо лучшие характеристики удается получить именно благодаря применению особо малых толщин резонатора [36], однако возможности технологии все же ограничивают минимально возможные толщины, поскольку толщину резонатора подбирают исходя из реально достижимых параметров поглощения и рассеяния света в материале [23]. Толщина резонатора Фабри—Перо должна составлять не менее половины длины волны света (для GaAs около $1/8$ мкм). Если бы материалы обладали исключительно высокой степенью нелинейности, тогда было бы выгодно использовать даже еще меньшую толщину резонатора, чтобы сократить время нарастания поля в резонаторе и чувствительность к температуре и длине волны. Следует заметить, что уменьшение поперечного размера до величины порядка длины волны представляет гораздо более трудную проблему по сравнению с получением минимальной толщины резонатора. При этом дифракционные потери света могут быть уменьшены за счет применения соответствующих волноводных структур, для изготовления которых, по-видимому, можно использовать метод травления [26], или, возможно, имплантацию. Хотя все это может быть непосредственно выполнено для активного материала, следует помнить, что создание структур

на основе резонаторов Фабри—Перо со специальными зеркалами, по всей видимости, приведет к значительному увеличению толщины и потребует применения специальных волноводных режимов работы устройства. В целом зеркала, созданные на основе четвертьволновых слоев AlGaAs с переменной концентрацией Al методом МЛЭ [37] или методом химического осаждения из паров металлоорганических соединений [38], могли бы изготавливаться тем же способом, что и объемные слои GaAs или квантоворазмерные структуры на GaAs в ОЛЭФП.

Резонаторы Фабри—Перо, использующие современные материалы при соответствующей оптимизации, способны работать в области частот 10—1000 ГГц. Большое быстродействие потребует либо гораздо более качественных материалов, чем имеющиеся в настоящее время, либо применения более сложных методик, которые усилят процесс релаксации. Отвод тепла от больших матриц с плотным размещением элементов, работающих в гигагерцовом диапазоне, может быть исключительно трудной задачей, однако методики охлаждения также совершенствуются. Матрица в 10^6 ячеек, потребляющая энергию в 1 фДж на ячейку на частоте в 1 ГГц, будет рассеивать мощность в 1 Вт. Промышленно освоенные охлаждающие микроэлементы [40] могут отводить до 1—2 Вт с площади в несколько квадратных миллиметров. Более того, эти охлаждающие элементы обеспечивают получение температуры до 77 К без применения жидкого азота, что дает возможность воспользоваться более «удачными» характеристиками полупроводников при этой температуре, чем при комнатной.

Различие в архитектуре создает определенные трудности при сравнительной оценке производительности вычислительных систем. В частности, мерой при оценке может служить число битов, обрабатываемых в секунду (бит/с). Так, самые быстрые компьютеры на сегодня обрабатывают около 10^{13} бит/с. Ожидается, что эта величина возрастет на один или два порядка, прежде чем будут реализованы оптические компьютеры. И хотя до сих пор были продемонстрированы только очень маленькие матрицы оптических элементов, переход к большим матрицам (по порядку величины до 10^6) не представляется слишком сложным. Были также продемонстрированы возможности получения скоростей далеко за пределами 1 ГГц. А основываясь на том предположении, что затраты энергии будут уменьшены до такой степени, что проблемы отвода тепла не будут стоять так остро, получим, что отдельная матрица может обрабатывать от 10^{15} до 10^{16} бит/с. Использование нескольких матриц, матриц большего размера или более быстродействующих устройств, может увеличить эту цифру даже на несколько порядков. Длительность оптического цикла составляет только 2 фс, поэтому в дальнейшем для оптических устройств до-

стижим терагерцевый диапазон. Вполне реальные возможности оптических вычислительных устройств должны, несомненно, стимулировать интенсивные исследования в этой области.

2.11. Три-стабильность и многозначная логика

Три-стабильность — это такая ситуация, в которой для одного и того же входного сигнала существуют три состояния выходного сигнала. Теория этого явления обсуждается несколькими авторами [40, 41] для случая двухфотонной системы. Оптическая три-стабильность поляризационных характеристик наблюдалась для паров натрия. В этой системе для линейно поляризованного падающего света прошедший свет имел три устойчивых состояния поляризации, а именно правая поляризация света с круговой поляризацией, левая поляризация света с круговой поляризацией и линейная поляризация. Соответственно с увеличением интенсивности входного сигнала переключается поляризация выходного сигнала.

Три-стабильность может быть использована для многозначной логики, как это было предложено в [42, 43]. Авторы предлагают использовать модифицированную цифровую систему с индексированными цифрами для выполнения «свободного от переноса разрядов» оптического сложения и вычитания. В этой цифровой системе используется набор цифр $\{\bar{1}, 0, 1\}$, где $\bar{1}$ — логическое дополнение к 1. Возможность чисто параллельного арифметического представления цифр в указанной системе может быть использована в три-стабильных системах.

2.12. Выводы

Почти все электронные компьютеры сегодня являются машинами последовательного типа. Это значит, что они выполняют относительно малое число операций за единицу времени. С другой стороны, оптическая информация легко передается параллельно и соответственно использование света в обработке информации, при принятии решений и в вычислениях может обеспечить намного большие скорости обработки данных. Параллельные оптические компьютеры позволят в перспективе одновременно выполнять миллионы операций, при этом одна операция будет требовать менее 1 нс. Это должно обеспечить скорости обработки, намного большие чем 10^{16} бит/с.

Шаги в направлении достижения этой цели уже предпринимались. Были созданы большие матрицы оптических логических элементов на GaAs [5, 26]. На рис. 2.9 показан фрагмент матрицы, состоящей из более чем 100×100 элементов, с размерами квадратной ячейки в 9×9 мкм. Изготовление матриц большего размера, например 1000×1000 элементов, не пред-

ставляется слишком сложным. Изучаются проблемы тепловыделения, каскадирования и создания параллельной архитектуры [44].

Исключительно важными являются исследования в направлении поиска оптических материалов с «быстрыми» и ярко выраженными механизмами нелинейностей. Целенаправленный подбор материалов с необходимыми свойствами дает возможность воспользоваться более удобными длинами волн, получить большее быстродействие и добиться работы приборов при повышенных температурах. Например, рост полупроводниковых многослойных квантоворазмерных структур с различными толщинами квантовых ям методом МЛЭ делает доступным экситонные резонансы на нужных длинах волн и энергиях связи; протонная бомбардировка снижает время жизни носителей, тем самым увеличивая быстродействие устройств. Полупроводники, особенно GaAs, являются наиболее многообещающими средами для нелинейной оптической обработки информации. Однако для некоторых применений, таких как волноводные устройства, могут стать пригодными органические и фоторефрактивные материалы.

КОМПОНЕНТЫ ОПТИЧЕСКОГО ПРОЦЕССОРА НА ОСНОВЕ ПРИБОРОВ С ЗАРЯДОВОЙ СВЯЗЬЮ (ПЗС)

*Б. Е. Бёрк, Лаборатория им. Линкольна,
Массачусетский технологический институт,
Лексингтон, шт. Массачусетс
Дж. Хиггинс, Центр исследований и разработки
микроэлектронных изделий, Роквелл
Интернэйшенэл, Саузенд Оукс, шт. Калифорния*

3.1. Введение

В настоящее время представляет большой интерес приложение оптических методов к обработке сигналов и вычислительным задачам. Скорость и параллельность обработки массивов данных, характерные для оптических методов, могли бы обеспечить значительное продвижение во многих задачах, требующих интенсивных вычислений. Поскольку системы, выполняющие «оптическую обработку сигналов» или «оптические вычисления», часто являются частью больших и в принципе электронных систем, то для обозначения этой конкретной области было бы более уместно использовать термин *электрооптическая обработка сигналов* (ЭООС). Стремительный рост быстродействия электронных устройств, особенно основанных на GaAs-полевых транзисторах, обеспечивает возможность объединения электронных и оптических процессоров для выполнения специальных сложных операций с недостижимыми ранее скоростями. Основными элементами, необходимыми для объединения электронной и оптической частей гибридного процессора, являются интерфейсные компоненты, соединяющие электронную и оптическую части. Именно в этой части технология создания ПЗС-приборов наиболее приспособлена для создания таких процессоров.

Одним из таких классов интерфейсных элементов являются детекторы света. Будучи объединенными с ПЗС-технологией, такие устройства могут выполнять гораздо более универсальные функции, чем просто преобразование света в электрические сигналы. Такие функции включают интегрирование, мультиплексирование, демультимплексирование и другие операции, необходимые для предобработки и задания форматов сигналов. Хотя кремниевая ПЗС-технология достигла впечатляющего уровня интеграции, формирователи изображений на арсениде галлия дают возможность получения намного больших скоростей.

Другой класс интерфейсных устройств представляют оптические модуляторы света. Пространственный модулятор света (ПМС) — это преобразователь, который накладывает изменяемую в пространстве картину модуляции света на падающий

пучок света. Такой шаблон определяется электрическими или оптическими сигналами, запрограммированными в устройстве. В данном конкретном случае рассматриваемый здесь элемент относится к типу поглощающих устройств (рис. 3.1), в котором излучение падает на поверхность двумерной матрицы элементов или ячеек. Оптическое пропускание каждой ячейки матрицы программируется независимо, так что свет, возникающий на выходе модулятора, имеет изменяющуюся в пространстве амплитуду. Фотопленочный транспарант является примером широко распространенного двумерного ПМС, но для будущих оптических процессоров наиболее необходимы модуляторы, способные быстро перепрограммироваться. Большие надежды на решение этой проблемы связывают с высоким быстродействием ПЗС на GaAs в сочетании с уникальными особенностями электропоглощения в составных полупроводниковых структурах.

В данной главе будут проанализированы приложения арсенид-галлиевой технологии к актуальным потребностям в быстродействующих матрицах фотоприемников и ПМС. Основное внимание уделяется особенностям, присущим лишь GaAs, в частности высокому быстродействию GaAs — ПЗС, и использованию явления электропоглощения в квантоворазмерных GaAs/AlGaAs-структурах. В разд. 3.2 рассматриваются некоторые примеры процессоров для ЭООС, использующие эти важнейшие интерфейсные компоненты. В разд. 3.3 предлагается обзор принципов работы ПЗС-структур на основе барьера Шоттки. В разд. 3.4 и 3.5 рассматриваются достижения в области создания быстро-

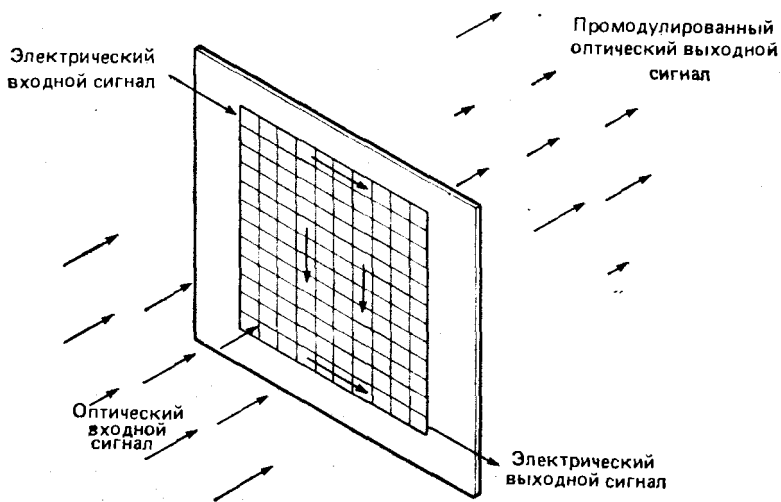


Рис. 3.1. Принципиальная схема двумерного пространственного модулятора света с адресацией на ПЗС.

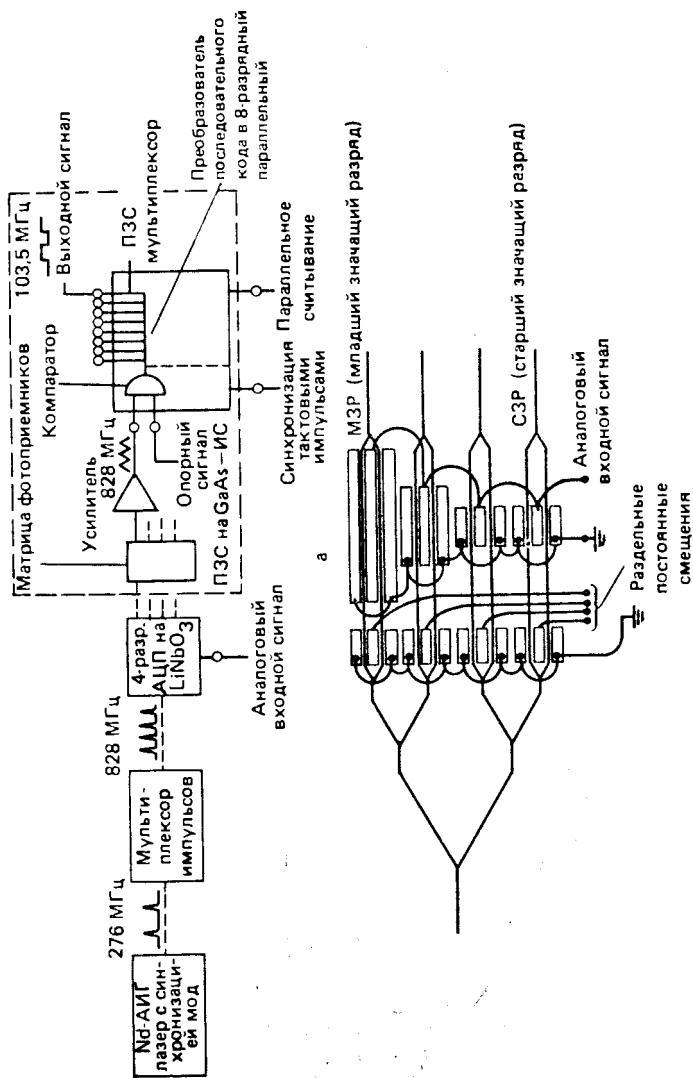
действующих ПЗС-матриц фотоприемников на GaAs и соответственно ПМС, основанных на эффекте электропоглощения в ПЗС.

3.2. Примеры электрооптической обработки сигналов (ЭООС)

Электрооптическая обработка сигналов обычно используется в задачах, где скорость и производительность более важны, чем точность и динамический диапазон. Это задачи из области обработки сигналов, обработки изображений и оптических вычислений. ЭООС особенно хорошо подходит для таких задач из-за изначально двумерной природы оптических пучков и возможности наложения нескольких пучков на одном и том же участке, что приводит к высокой степени параллельности вычислительных сетей. Эту концепцию параллельной обработки в настоящее время дублируют в электронных вычислительных сетях, называемых «систолическими матрицами», где большое число связанных процессоров образует упорядоченную двумерную сеть. В данном разделе представлены некоторые примеры такого параллелизма, откуда становится ясной превалирующая роль ПЗС-технологии.

Первым из примеров устройства ЭООС является аналого-цифровой преобразователь (АЦП), использующий электрооптический эффект в LiNbO_3 [3]. 4-разрядный вариант этого устройства показан на рис. 3.2, а, в то время как на рис. 3.2, б более подробно показана схема матрицы элементов АЦП на LiNbO_3 . Преобразуемый аналоговый сигнал подается на центральные электроды четырех оптических волноводных интерферометрических модуляторов, выполненных на подложке из LiNbO_3 . Эти модуляторы обладают тем свойством, что их оптический выходной сигнал представляет собой периодическую функцию аналогового сигнала напряжения и длины электродов вдоль направления волноводов. Это означает, что при подходящих пороговых характеристиках выходные сигналы волновода могут быть использованы для получения двоичного представления аналогового входного сигнала. Реализация функции выборки в этих устройствах осуществляется оптическими методами с использованием серии коротких лазерных импульсов, подаваемых с соответствующей скоростью выборки. Были получены скорости выборки в 1 ГВ/с (гигавыборка/секунду) [3], и методика дает надежду на получение скоростей, достигающих 3 ГВ/с.

На рис. 3.2, а пунктирной рамкой обозначена наиболее критическая часть преобразователя, регистрирующая короткие оптические импульсы, применяющая пороговую функцию к получаемому сигналу и затем демультимплексирующая данные



6

Рис. 3.2. 4-разрядный электрооптический аналого-цифровой преобразователь с быстродействующей регистрирующей системой на GaAs-элементной базе [3]:
 а — полная схема преобразователя; б — схематическое изображение матрицы АЦП на кристалле LiNbO_3 .

для дальнейшей обработки при меньших скоростях. В системе, предложенной в [3], эти функции осуществлялись с помощью лавинных фотодиодов, подсоединенных к быстродействующим биполярным компараторам и последовательно-параллельному преобразователю. При скоростях более 1 ГВ/с эти элементы схемы потребуются заменить на более быстродействующие, и здесь технология GaAs могла бы поднять существующий потолок быстродействия. Матричные ПЗС-детекторы на арсениде галлия могли бы обеспечить детектирование, а их выходные сигналы — усиливаться и подаваться в компараторы, изготовленные на монолитной GaAs-подложке. Демультимплексирование могло бы быть выполнено преобразователем последовательного кода в параллельный на ПЗС или быстродействующих полевых транзисторах.

Акустооптический спектроанализатор [17] представляет собой хорошо известный случай приложения оптических методов. На рис. 3.3 описана такая система и показано, как в ней используется ПЗС-детектор — демультимплексор. Анализируемый сигнал в ячейке Брэгга преобразуется в ультразвук. Лазерный луч при прохождении ячейки претерпевает дифракцию на бегущей акустической волне, и оптическая схема преобразует это угловое распределение в пространственное распределение в плоскости детектора. Таким образом, оптический сигнал в матрице детекторов представляет собой мгновенное фурье-преобразование радиочастотного сигнала.

Иллюстрацией больших возможностей параллельной оптической двумерной обработки служит показанная на рис. 3.4 схема вычисления разности в кадре. Основной операцией является сравнение элемент-за-элементом образа с ранее записанным в память образом. Записанный образ часто представляет собой более ранний образ того же самого объекта, и вычисление разности имеет целью обнаружение изменений объекта. Первый образ записан в двумерном ПМС с обратным контрастом. Таким образом, при прохождении второго образа через ПМС результатом будет исключение всей информации, кроме изменений, происходящих в образе. Хотя на рис. 3.4 описан «электрически» программируемый ПМС, окончательным устройством для таких применений будет служить «оптически» программируемый ПМС.

В заключение дан пример из области оптических вычислений. На рис. 3.5 описан метод умножения матрицы на матрицу, являющийся изящным примером оптического систолического процесса и названный архитектурой соединения матриц. Элементы двух матриц представлены ячейками оптических транспарантов, при этом пропускание света каждой ячейкой пропорционально элементам матрицы. Свет из однородного коллимированного источника слева проходит через транспаранты и ре-

гистрируется двумерной матрицей детекторов справа. Транспаранты одновременно смещаются перед детектором на 1 ячейку, как показано на рисунке, и при каждом шаге подается импульс света от источника. На рис. 3.5, а представлен момент времени, когда свет прошел через элементы транспарантов a_{11} и b_{11} , в результате чего свет с интенсивностью $a_{11}b_{11}$ собирается в одной ячейке матрицы фотоприемников. По мере выполнения

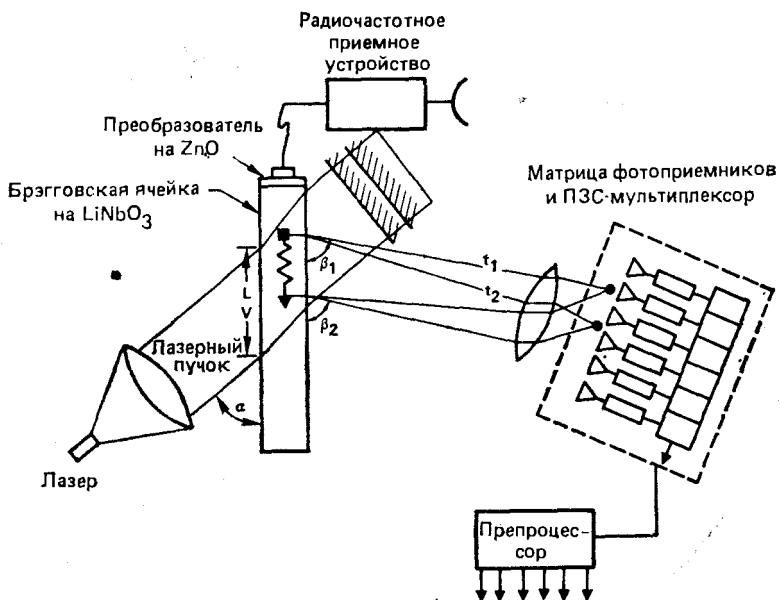


Рис. 3.3. Схема акустооптического спектроанализатора.

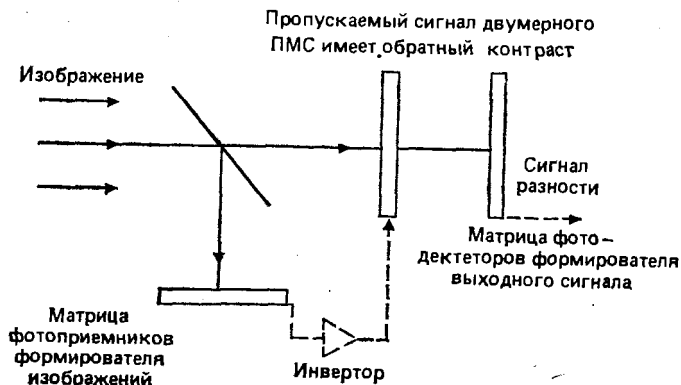


Рис. 3.4. Схематическое изображение метода получения разности двух изображений, в котором используются двумерные ПМС и матрицы фотоприемников.

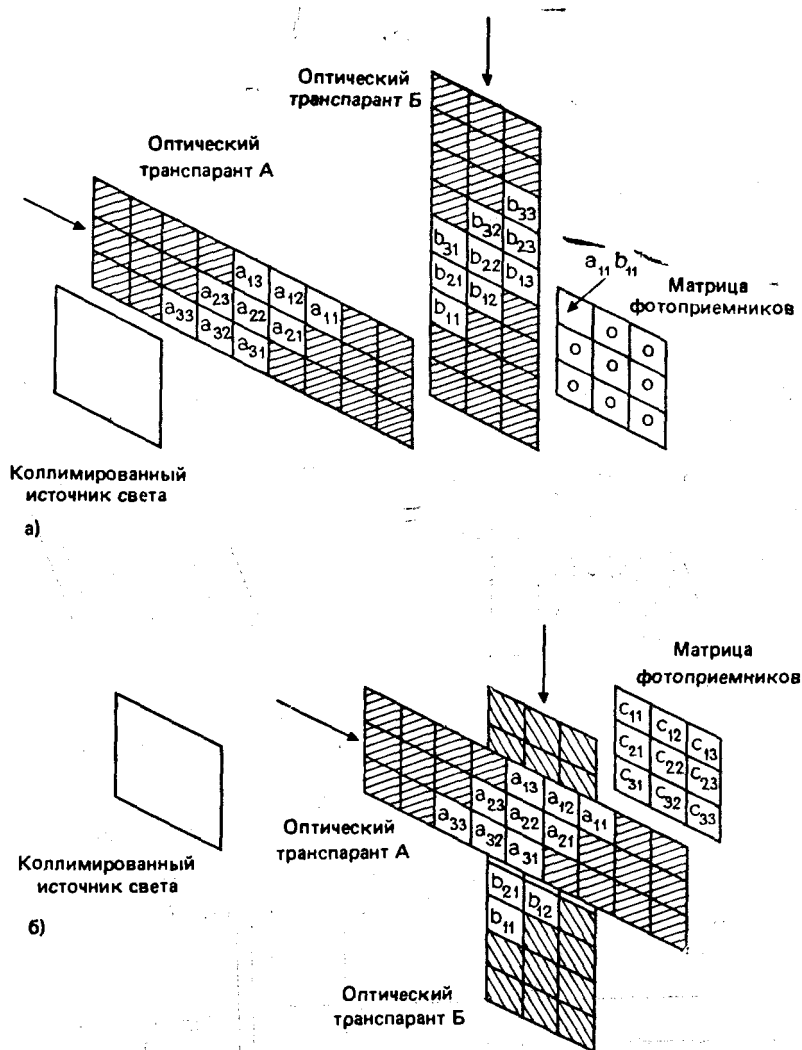


Рис. 3.5. Оптический умножитель матрицы на матрицу со встречными потоками данных, использующий «скользящие» оптические транспаранты [5]:
 a — исходное состояние;
 b — конечное состояние.

данного процесса детектор интегрирует суммируемые произведения элементов матриц. После перемещения транспарантов относительно детектора последний будет содержать требуемое произведение матриц (рис. 3.5, б). В дополнение к детектору в данном подходе требуется ПМС, обладающий способностью шагового перемещения или «прокрутки» значений пропускания относительно модулятора. В разд. 3.5 будет показано, что эта особенность является неотъемлемой чертой ПМС на основе ПЗС-структур.

Все эти примеры оптических процессоров имеют общий вид, показанный на рис. 3.6 и обычно включающий (1) источник света, (2) один или более пространственных модуляторов света, (3) оптические элементы (такие как линзы, зеркала, поляризаторы), (4) матрицу детекторов. Отсюда следует, что модуляторы и детекторы являются устройствами, где применение ПЗС-технологии может внести значительный вклад, и в оставшейся части этой главы обсуждаются эти устройства.

3.3. Принципы работы ПЗС

Чтобы полностью понять возможности GaAs-ПЗС в области систем оптической обработки сигналов, необходимо познакомиться с физикой этих приборов. В данном разделе дается описание устройства и принципов работы ПЗС на основе барьеров Шоттки. Обсуждение будет по необходимости коротким, и основное внимание уделено аспектам, наиболее характерным проблеме оптической обработки с большой скоростью. Приборы с зарядовой связью были рассмотрены исчерпывающим образом в [25], а недавние разработки в области ПЗС на основе барьеров Шоттки обсуждались в [24]. В дополнение к GaAs будет упомянут и важный родственный материал AlGaAs, который непосредственно составляет в комбинации с GaAs высококачественные гетеропереходы и, похоже, будет иметь большое значение в устройствах ЭООС.

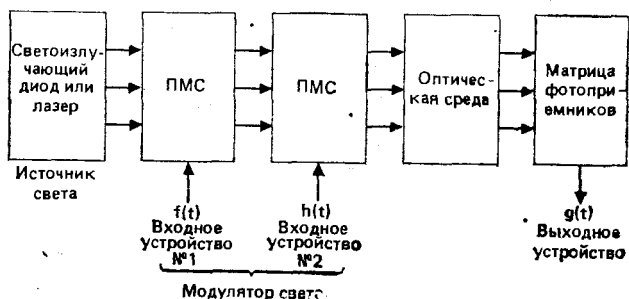


Рис. 3.6. Принципиальная схема электрооптического процессора.

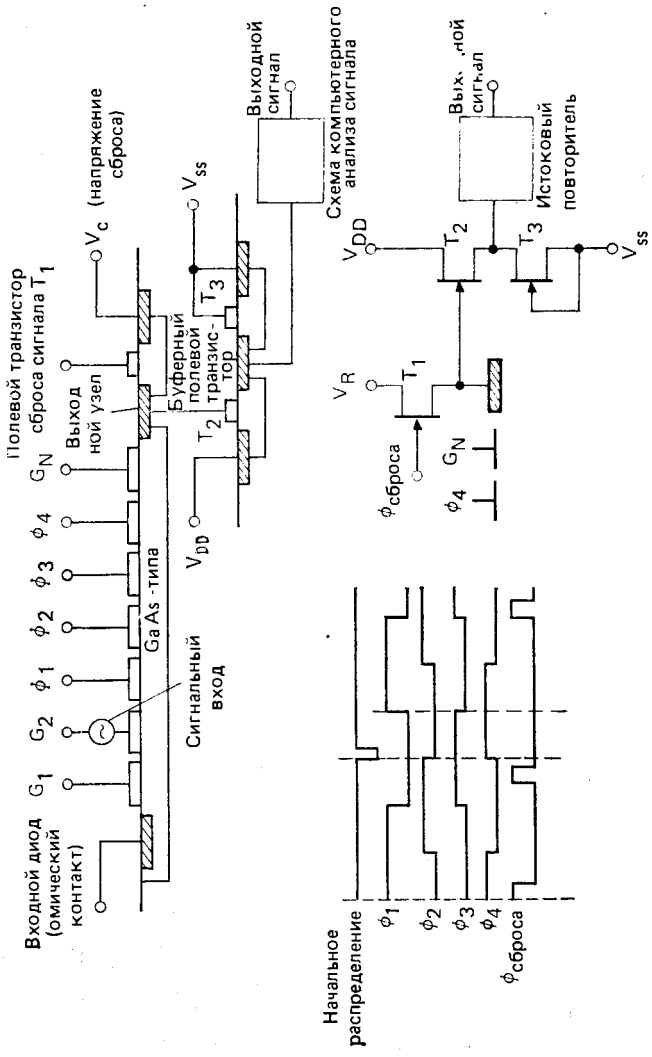


Рис. 3.7. Поперечное сечение 4-фазной ПЗС-структуры на основе барьеров Шоттки с выходной регистрирующей схемой на полевых транзисторах.

На рис. 3.7 показан разрез GaAs ПЗС-структуры со скрытым каналом, сделанным на основе барьеров Шоттки. Термин «скрытый канал» относится к тому обстоятельству, что зарядовые пакеты удерживаются в n -легированном слое или канале, что отличает это устройство от приборов с поверхностным каналом, в которых заряд скапливается в поверхностном инверсном слое МОП или МДП-структуры. Существенными чертами являются омический входной контакт, через который носители инжектируются в устройство, матрица близко расположенных затворов на основе барьеров Шоттки и выходной омический контакт (обозначенный «выходной узел»), который принимает с временной задержкой заряд сигнала. Устройство управляется серией сигналов, форма которых зависит от времени; примеры некоторых из них показаны на рисунке. Назначение этих «тактовых» сигналов состоит в том, чтобы произвести выборку входного сигнала (подаваемого на G_2 в данном варианте схемы),

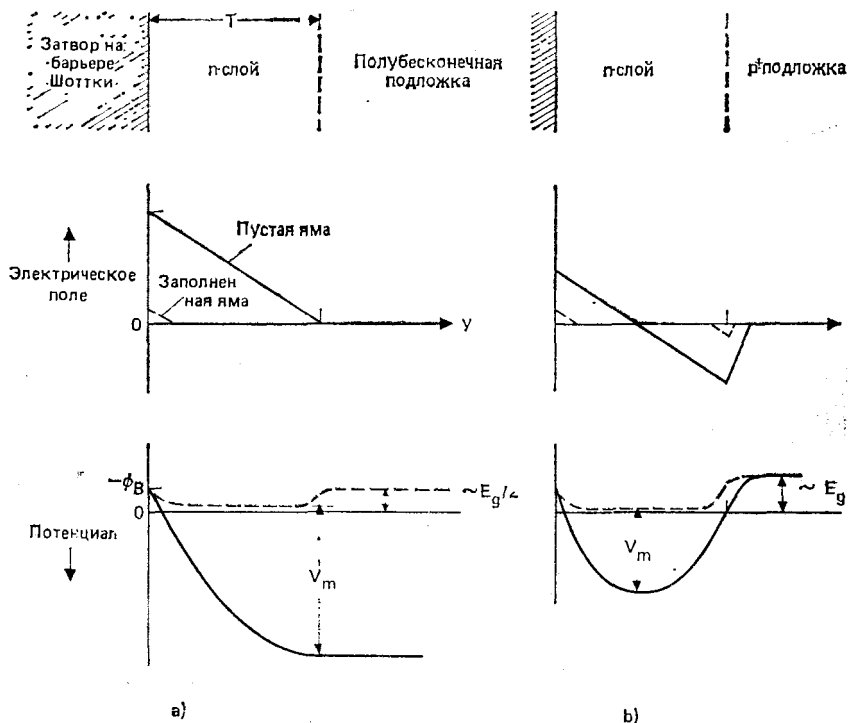


Рис. 3.8. Распределение потенциала и электрических полей в ПЗС на барьерах Шоттки.

Случай а) относится к более распространенному варианту размещения n -слоя на полуизолирующей подложке, в то время как б) относится к однородно легированному n -слою на r^+ -подложке.

преобразовать его в зарядовый пакет, пропорциональный амплитуде сигнала, и переместить этот пакет на выходной узел, управляя этим процессом с помощью электродов у соответствующих затворов. Вспомогательные цепи, показанные на рис. 3.7, обычно размещаются на той же подложке, включая полевые транзисторы затвора сброса и истоковый повторитель, используемый для обнаружения заряда на выходном узле. В некоторых случаях формирователи тактовых импульсов и дополнительные цепи согласования выходного сигнала также выполняются интегрально с ПЗС.

Чтобы понять, как возникает перенос заряда, на рис. 3.8 и 3.9 показаны электростатические потенциалы и электрические поля в материале. Потенциалы отсчитываются относительно дна зоны проводимости, потенциал затвора берется за начало отсчета при 0 В. На рис. 3.8 изображены распределения потенциала и электрических полей вдоль оси y (по нормали к плоскости устройства) для двух случаев. В случае (а) слой GaAs состоит из однородно легированного слоя n -типа на полуизолирующей подложке, в то время как случай (б) дает описание слоя n -типа

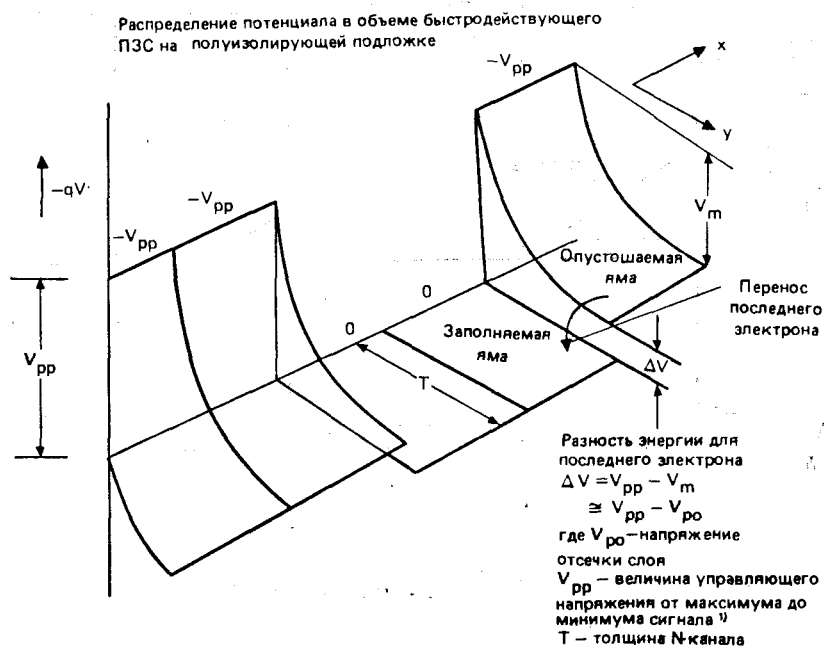


Рис. 3.9. Объемное распределение потенциала в канале ПЗС. Заряд переносится в направлении x параллельно плоскости устройства, соответствующей xz -плоскости.

1) Данная величина представляет собой удвоенную амплитуду и называется размахом,

на подложке p^+ -типа. Случай (а) представляет существенно более общую ситуацию, поскольку меньшей является емкостная нагрузка, создаваемая подложкой, и последующее обсуждение относится к этому случаю. Сплошные кривые относятся к условию пустых ям, т. е. электроны n -слоя полностью вырождены. Пунктирная линия соответствует случаю заполненной ямы, где электроны почти полностью нейтрализуют заряд доноров, за исключением узких областей вблизи затвора и границ раздела подложки. В обоих случаях в n -слое возникает максимум потенциала (или минимум энергии электронов). Разность V_m между потенциалом пустой ямы и потенциалом полной ямы может быть задана простым выражением для случая (б):

$$V_m = V_{po} - V_{bi}, \quad (3.1)$$

где $V_{bi} = \phi_b - (kT/q) \ln(N_c/N)$ — контактная разность потенциалов и ϕ_b — высота барьера на границе металл-диэлектрик, N — концентрация легирования донорами, N_c — плотность состояний в зоне проводимости ($4,7 \times 10^{17}$ см $^{-3}$ для GaAs) и kT/q — напряжение, соответствующее тепловой энергии. Напряжение отсечки V_{po} задается выражением

$$V_{po} = \left(\frac{qNT^2}{2\epsilon} \right)^{1/2}, \quad (3.2)$$

где T — толщина слоя, q — заряд электрона, ϵ — диэлектрическая проницаемость полупроводника. Во всех случаях, представляющих практический интерес, $V_{bi} < 1$ В и $V_m \approx V_{po}$.

В направлении x (направление переноса заряда) поля и потенциалы определяются в основном напряжением на затворе. На рис. 3.9 представлены трехмерные зависимости распределения потенциалов в канале ПЗС для случая n -слоя на полуизолированной подложке. Два затвора находятся под нулевым напряжением и были заполнены зарядом, перенесенным из смежных ям, находящихся под напряжением $-V_{pp}$. Из этой диаграммы ясно, что, для того чтобы полностью очистить «опустошаемую» яму, размах (удвоенная амплитуда) напряжения тактовых импульсов V_{pp} должен быть больше, чем максимальное напряжение в канале V_m . Практически необходим размах тактового сигнала около $1,5 V_m$, чтобы обеспечить полный перенос электронов в соседнюю яму.

Управляющие сигналы, прикладываемые к затворам, показаны более подробно на рис. 3.10 для случая четырехфазной синхронизации устройства. При низких частотах тактовые сигналы являются прямоугольными сигналами, распространяющимися между высоким потенциалом, который притягивает электроны в потенциальную яму, находящуюся под затвором, и низким потенциалом, отталкивающим электроны. В любой заданный момент времени по крайней мере два соседних затвора

находятся под высоким потенциалом и накапливают заряд. Для перемещения зарядового пакета на следующую пару затворов потенциал на третьем затворе (например, ϕ_3) должен расти, в то время как потенциал первого затвора (например, ϕ_1) уменьшается. Тогда заряд выталкивается потенциалом ϕ_1 из опустошаемой ямы, в области с ϕ_2 и ϕ_3 , где ямы заполняются.

По мере увеличения тактовой частоты становится более трудным поддержание быстрых тактовых переходов, и форма импульсов ухудшается почти до синусоидальной формы. К счастью, большинством ПЗС-структур можно управлять синусоидальными сигналами, без значительного снижения характеристик. Этот факт может быть использован для снижения рассеиваемой мощности в цепях формирователей тактовых сигналов. Причина этого состоит в том, что затворы ПЗС-структур представляют преимущественно емкостную нагрузку для цепи формирователя и, следовательно, возможно управлять затворами с помощью RC-цепей, обладающих высокой добротностью. В этом случае потребление мощности может быть намного ниже, чем для традиционных переключающих цепей, рассеивающих существенные мощности.

Одну важную проблему, связанную с этим устройством, не указанную на слишком упрощенной диаграмме на рис. 3.9, представляет область зазора между затворами, показанная на рис. 3.7. Если этот зазор очень широкий, соседние электроды на затворах не регулируют соответствующим образом потенциал в канале и возможно образование ложных потенциальных ям или барьеров, которые в свою очередь могут вызвать серьезные нарушения в процессе переноса заряда [8, 9]. Чтобы избежать этой опасности, требуется поддерживать минимальные размеры зазора, обычно не более 1 мкм. Это является особенно

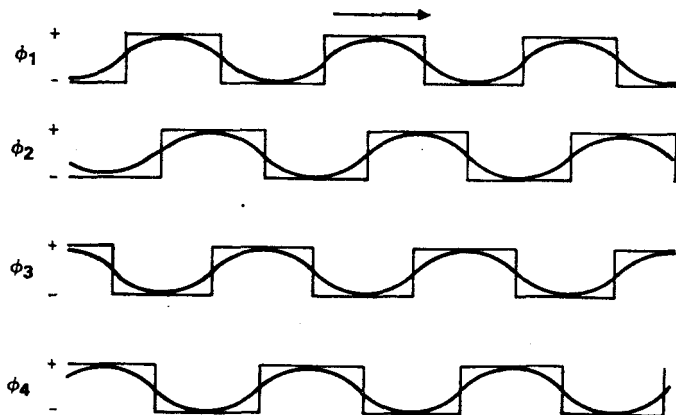


Рис. 3.10. Формы тактовых сигналов для 4-фазного ПЗС.

настоятельным требованием в процессе изготовления устройства и нуждается в совершенном литографическом оборудовании и тщательном проведении процессов. Недавно была продемонстрирована новая структура с затворами, изготавливаемыми из одного или более металлических слоев и диэлектрических слоев, обеспечивающих межэлектродную изоляцию [20]. Этот процесс создания перекрывающихся затворов предъявляет менее жесткие требования к литографии и дает возможность регулировать размеры зазора с субмикронной точностью. Ниже описана другая структура затворного типа, успешно работавшая с гигагерцевыми тактовыми частотами.

Далее будут кратко упомянуты ограничения характеристик, накладываемые темновым током и ловушечными эффектами. Темновой ток представляет собой спонтанную генерацию электрон-дырочных пар и, следовательно, является источником нежелательного заряда. Так как представляющие здесь интерес приложения связаны с использованием высоких тактовых частот, заряд темнового тока обычно будет удаляться из устройства прежде, чем он будет накоплен в значительном количестве. Типичная величина для GaAs составляет 100 нА/см^2 , что приблизительно в 100 раз больше, чем для существующих кремниевых ПЗС. На этом уровне типичная ПЗС-структура на GaAs будет заполнена темновым током за времена около 0,1 с. AlGaAs имеет более широкую запрещенную зону, чем Si и GaAs, и вследствие этого могут быть получены намного более слабые темновые токи. При комнатной температуре в ПЗС-структурах на AlGaAs наблюдались значения, близкие к $0,043 \text{ нА/см}^2$ [18].

Даже когда ПЗС-структура работает при достаточном размахе тактовых импульсов и имеет хорошие профили распределения потенциала в области между затворами, обеспечивающими полный перенос свободных зарядов, остается дополнительный источник потерь заряда, возникающий вследствие захвата на ловушки. Источником таких ловушек являются химические загрязнения и структурные дефекты кристалла. Захват носителей на такие состояния обычно происходит за времена в пределах наносекунд, но освобождаются они за времена, зависящие от типа ловушки. По сравнению с кремнием, где обычными являются концентрации ловушек менее 10^{11} см^{-3} , плотности ловушек в GaAs и других смешанных полупроводниках обычно составляют по меньшей мере в 100 и более раз большие величины. Но, несмотря на это, сообщалось о создании ПЗС-структур на GaAs с эффективностью переноса, приближающейся к кремниевым устройствам [26]. Одна из причин может состоять в том, что скорость освобождения заряда или время эмиссии из ловушки изменяются в очень широких пределах и, следовательно, различные ловушки будут обладать различными степенями влияния на характеристики устройства. Например,

один особенно широко распространенный вид ловушек в GaAs — это дефект стехиометрии, известный как EL2, время эмиссии которого при комнатной температуре составляет около 10 с. Ясно, что ловушка этого типа не будет оказывать влияние на работу устройства в такой же значительной степени, как ловушка, имеющая время эмиссии порядка периода тактового цикла. В настоящее время природа ловушечных эффектов в ПЗС-структурах на GaAs в основном не исследована.

3.3.1. Обсуждение конструкций с высоким быстродействием

Ясно, что при больших скоростях работы носители в канале ПЗС-структур должны испытывать влияние значительной компоненты поля E , параллельной направлению переноса. Численные расчеты показывают, что максимум поля, переносящего заряд, возникает на расстоянии около $0,4L$ в глубь канала, где L — длина затвора [25]. Наоборот, это поле стремится иметь низкие значения под центром затвора вблизи поверхности полупроводника из-за закорачивающего действия металла. Это предполагает, что при работе с большими скоростями канальный слой должен быть довольно толстым, обычно микрон или более. Однако использование толстого слоя вступает в противоречие с двумя другими аспектами конструкции устройства. Во-первых, толстый слой будет иметь в соответствии с уравнением (3.2) высокое напряжение отсечки, и это сделает необходимым соответственно высокий размах тактовых напряжений. Напряжение отсечки может быть уменьшено при снижении N , но за счет приносимой в жертву емкости, определяемой зарядом (пропорциональной NT) и, следовательно, динамического диапазона. Вторая проблема состоит в том, что для работы с большой скоростью размещенные на чипе вспомогательные цепи, такие как выходные полевые транзисторы или формирователи тактовых импульсов, требуют применения тонких канальных слоев. На рис. 3.11 изображены эти противоречивые требования к n -слою с концентрацией доноров N на полубесконечной подложке. Кривая B — это линия постоянного напряжения отсечки, составляющего 3,5 В. Это значение выбрано потому, что для применений при гигагерцевых тактовых частотах максимальный размах тактового напряжения не должен выходить за пределы от 5 до 7 В. Кривые постоянной, определяемой зарядом емкости NT , показаны пунктирными линиями, и ясно, что динамический диапазон быстро уменьшается для толстых слоев. Участок, обозначенный A , однако, является типичной областью параметров для конструкций полевых транзисторов на широкозонном GaAs и, следовательно, является желательной областью режимов для вспомогательных электронных цепей на полевых транзисторах. Таким образом, требования к ПЗС-струк-

турам и полевым транзисторам взаимно удовлетворены для тонких сильнолегированных слоев. Проблема с областью A состоит в том, что слои указанной толщины требуют использования длин затворов ПЗС-структур, нереально малых для работы при гигагерцевых тактовых частотах.

Для одновременного удовлетворения требований к материалу, заданных областью A на рис. 3.11, и обеспечения высоких тактовых скоростей без использования чрезмерно коротких затворных площадок была разработана структура затворов, которая отличается от вышеуказанной. Эта структура называется ПЗС-структурой с резистивными затворами и приведена на рис. 3.12 [13]. Затворные электроды состоят из узких металлических полосок, разделенных сравнительно широкими промежутками с нанесенной сверху дополнительной резистивной пленкой. Резистивная пленка должна удовлетворять тем же требованиям, что и металл, а именно образовывать хороший барьер Шоттки с полупроводником. Подходящие материалы включают керметы и силниды тугоплавких металлов. Цель этого подхода состоит в том, чтобы гарантировать наличие сильного горизонтального электрического поля вблизи поверхности. В дополнение этот метод позволяет избежать неконтролируемых полей в канале в междуэлектродных зазорах, характерных для описанных выше традиционных структур.

На рис. 3.13 показаны расчетные зависимости времен, необходимых для переноса заряда с эффективностью в 99,99%, от толщины T n -слоя. Представлены данные для двух структур с проводящими затворами и одной структуры с резистивными затворами. Заштрихованная область указывает диапазон величин T , желательный для обеспечения работы полевых транзисторов на высоких частотах. Видно, что устройства с резистивным затвором имеют лучшие характеристики, особенно при меньших толщинах слоя. Заметим, что четырехфазное устройство с $T=0,25$ мкм могло бы иметь время переноса менее 250 пс для тактовых частот более 1 ГГц. Эта цифра также учитывает фактор заполнения Q_F , представляющий отношение фактической емкости заряда, деленной на полный заряд донорных центров под двумя затворами, и показывающий, что устройства с резистивным затвором также имеют более высокую емкость заряда для тонких слоев. Недавно полученные экспериментальные результаты для этих устройств дали значение эффективности переноса заряда более 0,99 при 2,5 ГГц [27].

3.4. Применение ПЗС в качестве выходных устройств ЭООС

В разд. 3.2 было указано, что в системах ЭООС выходными устройствами являлись детекторы, роль которых состояла в преобразовании оптических сигналов в эквивалентные электри-

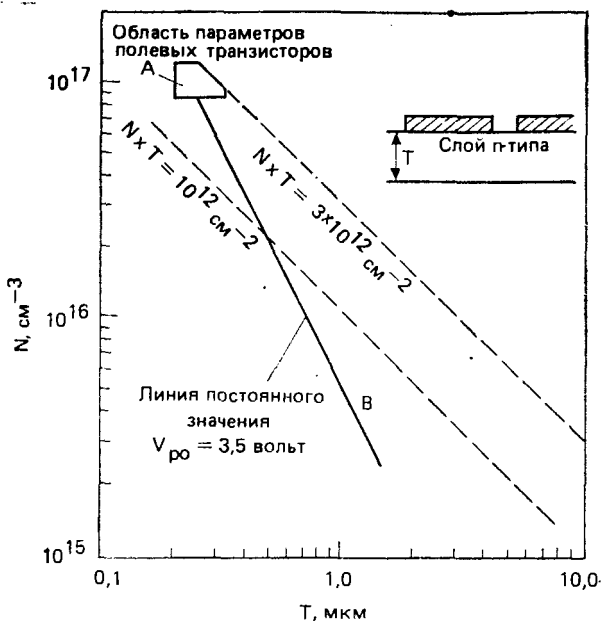


Рис. 3.11. Соотношение между уровнем легирования N и толщиной T в слое n -типа на полубесконечной подложке быстродействующей ПЗС-структуры (линия В) и область широкополосного режима работы полевых транзисторов (область А).

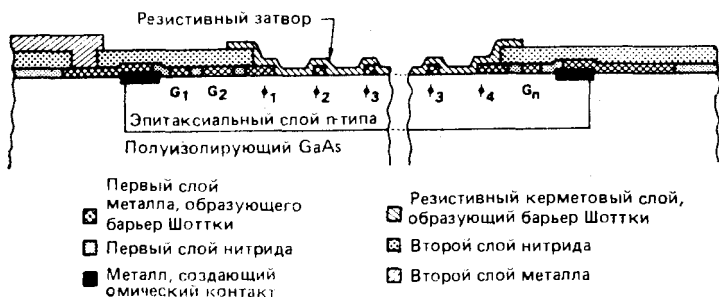


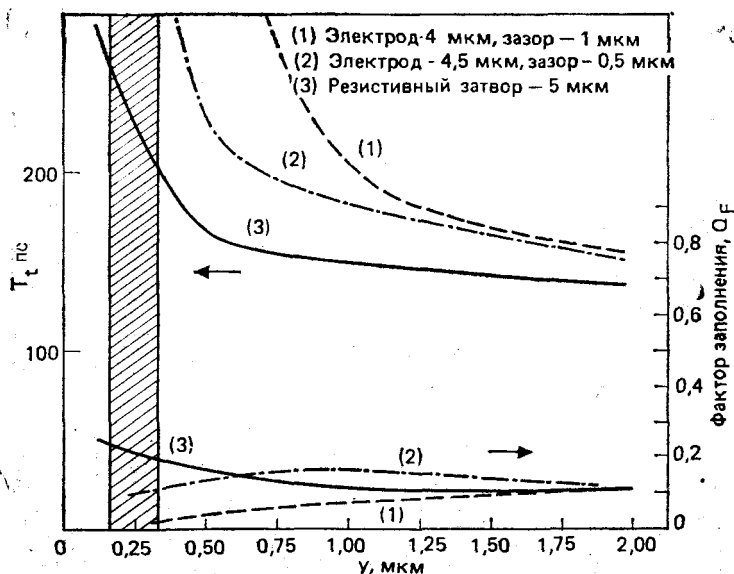
Рис. 3.12. Поперечное сечение ПЗС-структуры с резистивным затворным слоем.

ческие. Свойства ПЗС-структур хорошо удовлетворяют не только процессу детектирования, но также другим вспомогательным функциям, таким как интегрирование по времени и мультиплексирование. В данном разделе описаны некоторые примеры ПЗС-структур на GaAs, которые реализуют эти функции.

Хотя развитие кремниевых детекторов и формирователей изображений на ПЗС-структурах находится на совершенном уровне, эти устройства для некоторых применений не обеспечи-

вают требуемую ширину полосы частот. Примером этого является матрица детекторов для акустооптического спектроанализатора, показанная на рис. 3.3. Данная система часто используется для измерений параметров импульсных и других быстроизменяющихся сигналов с широким диапазоном, и такие приложения требуют предельных скоростей детекторов и последующих процессоров. На рис. 3.14 показаны принципиальные особенности матрицы детекторов и мультиплексора. Рассеянное в брэгговской ячейке лазерное излучение падает на матрицу фотодетекторов (сектор А), выходное напряжение или ток которой проходит в схему, обозначенную «интегратор/уплотнитель сигнала» (сектор Б). Так как динамический диапазон входного оптического сигнала в таких системах очень велик, эта цепь, если необходимо, могла бы выполнять нелинейное преобразование сигнала детектора в заряд, чтобы сжать динамический диапазон до уровня, совместимого с ПЗС. Тогда заряд следовало бы проинтегрировать в заданном интервале времени и параллельно ввести в ПЗС-мультиплексор для считывания (сектор В).

На рис. 3.15 показана фотография матрицы на 32 детектора с буферным услителем, выходной сигнал которого мультиплексируется в выходной регистр ПЗС-структуры [23]. ПЗС-муль-



Определение y : координата движущегося электрона, отсчитываемая в глубину электрода

Рис. 3.13. Расчетные зависимости времени переноса заряда T_t и фактора заполнения Q_F от толщины n -слоя.

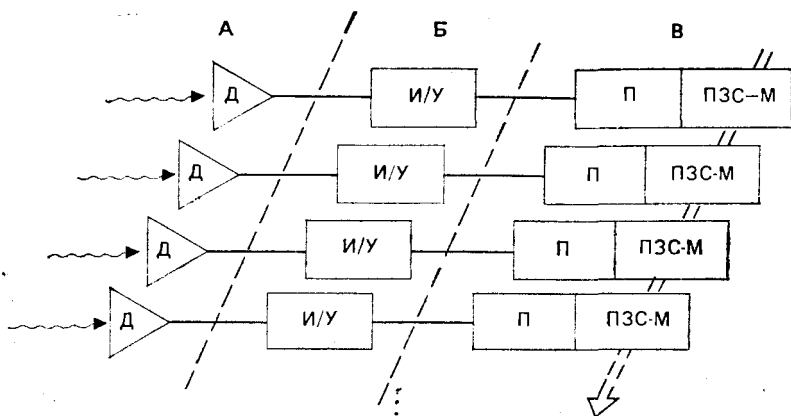


Схема генерации сигнала, включающая матрицу фотодетекторов и ПЗС-мультиплексор

А. Матрица детекторов

Б. Интегратор/уплотнитель сигнала (индивидуально настраиваемый)

В. Параллельный ПЗС-мультиплексор на полевых транзисторах

Г. Параллельная система ввода в ПЗС

Рис. 3.14. Схема матрицы фотоприемников и мультиплексора, предназначенных для использования в акустооптическом спектроанализаторе.

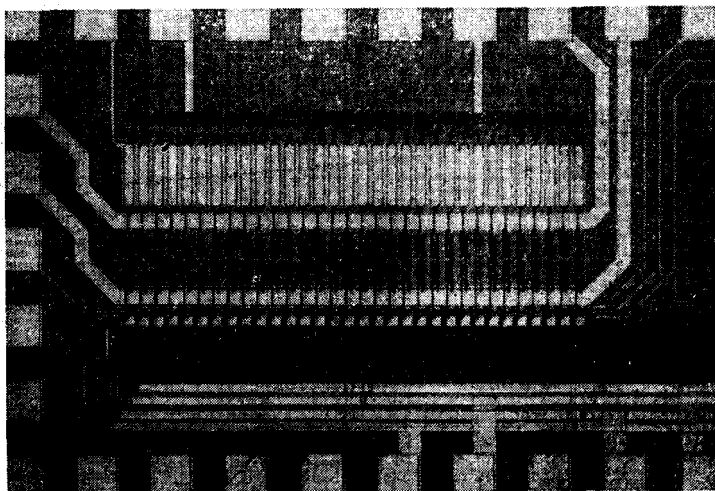


Рис. 3.15. Фотография ПЗС-детектора/мультиплексора на GaAs. Линейка из 32 детекторов и буферных усилителей (в центре) подает сигналы в 4-фазный сдвиговый регистр ПЗС (внизу).

типлексор с параллельным боковым вводом продемонстрировал работу с тактовой частотой 1 ГГц и эффективностью переноса заряда более 0,997.

Другим примером, где может оказаться полезным повышенное быстродействие GaAs-технологии, является обработка некогерентных оптических сигналов, изображенная на рис. 3.16 [22]. Этот процессор выполняет умножение вектора и матрицы, используя простую электрооптическую методику. Вектор f представлен временной последовательностью сигналов, модулирующих светодиод. Сигнал от светодиода проходит через маску, состоящую из апертур, площадь которых соответствует величине матричных элементов h_{mn} и собирается ПЗС-формирователем изображений, ячейки которого установлены на одной прямой с изображениями апертур. Свет, собранный в ячейке (m, n) вследствие прохождения светового импульса, соответствующего элементу вектора f_n , представляет произведение f_n и матричного элемента h_{mn} . Для каждого нового образца f_n зарегистрированный заряд в ПЗС должен быть сдвинут вправо на один элемент, и по завершении процесса результирующая матрица произведения g_{mk} считывается выходным регистром. Для более высоких выходных скоростей мультиплексирование внешним регистром было бы исключено и выходные сигналы брались бы прямо из каждой строки.

Пример устройства, использующего одномерную ПЗС-структуру на GaAs, был описан в [13]. ПЗС-структура, показанная сверху на рис. 3.17, имела прозрачные электроды затвора и маску с модулированной апертурой, изготовленную интегрально непосредственно сверху на электродах затворов. Результаты тестирования, показанные на двух нижних фотографиях, были получены при освещении устройства короткими лазерными импульсами и считывании заряда со скоростью 1 МГц. Выходной сигнал ПЗС-структуры (нижнее фото) повторяет апертуру входного сигнала, которая видна на фотографии устройства сверху на рис. 3.17. Электрический входной сигнал, состоящий из двух импульсных выбросов (среднее фото), подтверждает, что эффективность переноса заряда является высокой.

Двумерный вариант такого устройства способен выполнять разнообразные линейные преобразования, такие как корреляция, фильтрация и фурье-преобразование, как было показано для кремниевых формирователей изображения. Для структур на GaAs такие операции несомненно могли выполняться при ширине полосы частот, приближающейся к 1 ГГц. Хотя возможности расширения полосы частот ПЗС-структур на GaAs были хорошо изучены, технология получения двумерных формирователей изображения отработана плохо. Некоторые начальные результаты, полученные с устройствами в 32×32 элемента, являются обнадеживающими [18].

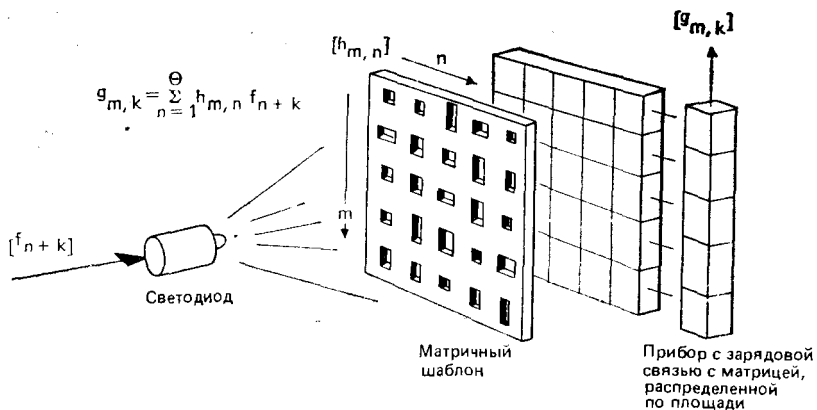


Рис. 3.16. Принципиальная схема электрооптического процессора [22].

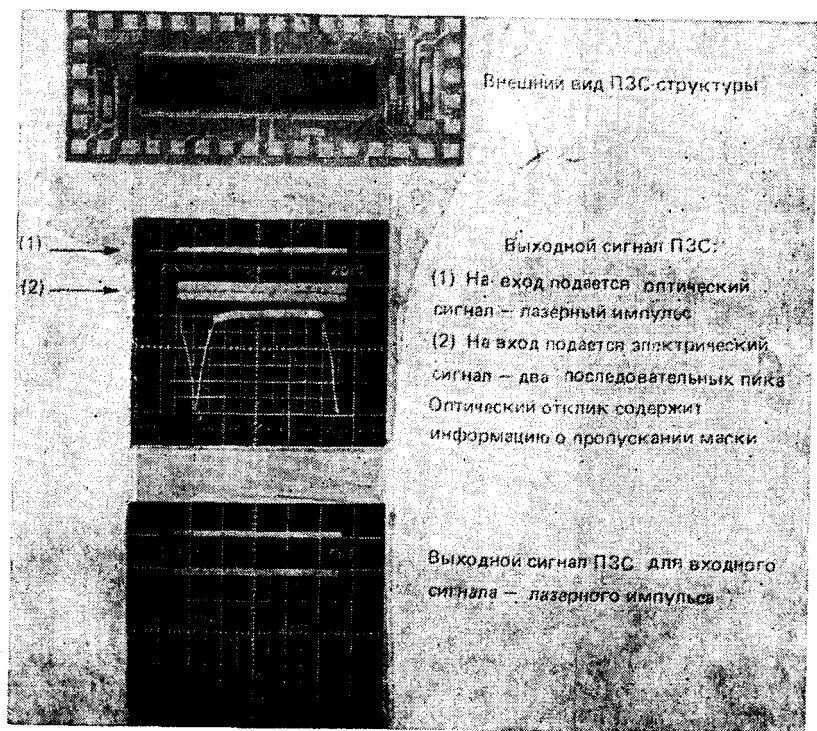


Рис. 3.17. Общий вид и характеристики электрооптического процессора, выполненного по схеме, показанной на рис. 3.16, и использующего ПДС-структуру с затеняющей маской, изготовленной интегрально с прибором.

Единственное заключительное замечание, представляющее особый интерес, состоит в том, что ПЗС на GaAs могут работать при криогенных температурах. Экспериментальные данные по ПЗС-структурам на GaAs показывают, что для тактовых частот в 10 МГц эффективность переноса заряда начинает спадать при температуре около 14 К. В отличие от этого эффективность переноса кремниевых ПЗС со скрытым каналом начинает ухудшаться при температурах ниже 70 К. При этих температурах быстроедействие ПЗС на GaAs, как ожидается, будет по меньшей мере таким же хорошим и, вероятно, даже более высоким, чем при комнатной температуре. Эти проявляемые при низких температурах свойства оказываются полезными в гибридных инфракрасных фокусируемых матрицах детекторов, где ПЗС состыковывается с длинноволновыми детекторами, такими как HgCdTe. Кроме того, коэффициенты теплопроводности GaAs и HgCdTe хорошо согласованы между собой. Рассеяние мощности представляет другую важную проблему в таких фокусируемых матрицах вследствие работы при криогенных температурах. Однако ПЗС рассеивают малую мощность потому, что как упомянуто выше, они представляют в основном емкостную нагрузку для цепей формирователей тактовых импульсов.

3.5. Адресуемые ПЗС-пространственные модуляторы света (ПЗС-ПМС)

Отсутствие подходящих ПМС является, вероятно, наиболее серьезной помехой для быстрого развития обработки оптических сигналов и данных. Важными рабочими характеристиками таких устройств являются высокая скорость записи, высокое разрешение (большое число элементов), большая глубина модуляции. Сочетание этих черт в одном устройстве было недостижимой целью, даже при том, что существует большое число концепций, выдвинутых на различных стадиях развития тематики [10, 30].

В этом разделе мы описываем современное состояние исследований, направленных на соединение явления электропоглощения и ПЗС-технологии, с целью получения быстродействующих модуляторов с высоким разрешением и высоким контрастом. В таком устройстве зарядовый пакет в потенциальной яме в ПЗС изменяет электрическое поле и, следовательно, оптическое пропускание электропоглощающей среды. В представленном на рис. 3.1 ПЗС-устройстве зарядовый пакет может быть электрически введен во входной регистр (он показан вдоль верхней части устройства), и затем перенесен параллельно регистрам столбцов. Тогда заряд может быть перемещен серией тактовых циклов вниз по столбцам, и когда все ячейки заполнены, то может быть подан оптический входной сигнал, который требуется

промодулировать. Эта возможность шагового смещения или «прокрутки» данных в устройстве является привлекательной чертой для определенных концепций оптических вычислений, таких как схема RUBIC¹⁾ [5]. В качестве альтернативы ПЗС может работать как формирователь изображений, так что целый шаблон для модуляции света может быть установлен оптически за времена менее наносекунд. Для емкости ямы в ПЗС, составляющей 10^6 электронов, и квантовой эффективности в 50% требуемая оптическая мощность для длины волны 850 нм составляет 0,5 пДж.

3.5.1. Электропоглощение

Монолитный ПМС на GaAs ПЗС-структуре основан на явлении электропоглощения, или разновидности оптического поглощения в электрическом поле. Будут описаны два примера электропоглощения, подходящие для реальных приборов. Первый — это эффект Франца—Келдыша, в котором оптическое поглощение при энергиях фотонов, немного меньших, чем ширина запрещенной зоны, может быть увеличено при приложении электрического поля. Второй случай, обсуждаемый ниже более подробно, представляет недавно обнаруженный сдвиг экситонного поглощения в квантовых ямах на GaAs/AlGaAs и представляющий особый интерес из-за того, что изменение поглощения во много раз сильнее, чем в случае эффекта Франца—Келдыша. Оба эффекта имеют необходимое быстродействие, так как они зависят от квантовомеханических явлений, возникающих в субпикосекундных временных масштабах. На практике быстродействие модуляторов, использующих эти эффекты, будет ограничено процессом переноса заряда в ПЗС.

Эффект Франца—Келдыша заключается в изменении оптического поглощения в электрическом поле для фотонов с энергией, близкой к ширине запрещенной зоны (данному эффекту посвящены обзоры [1, 4]). Этот эффект можно наблюдать в GaAs и во всех прямозонных полупроводниках, но, к сожалению, он пренебрежимо мал в непрямозонных материалах, таких как кремний. Теоретические и экспериментальные зависимости коэффициента оптического поглощения от длины волны при различных напряженностях электрического поля показаны на рис. 3.18 для GaAs [28] и четверных полупроводниковых соединений $\text{Ga}_{0.24}\text{In}_{0.76}\text{As}_{0.52}\text{P}_{0.48}$ [14]. Единственные параметры материала, используемые в выражениях для электропоглощения, это эффективная масса и показатель преломления. Поскольку

¹⁾ Полное название схемы RUBIC cube переводится как «быстродействующее некогерентное биполярное вычислительное устройство на светоделительном кубе, работающее без напряжения». — Прим. перев.

эти параметры приблизительно одинаковы для GaAs и GaInAsP, то расчетные кривые на рис. 3.18 должны быть применимы к обоим материалам. Для обоих полупроводников измеренные величины электропоглощения оказываются несколько ниже, чем следует из результатов теоретических расчетов. Маловероятно, что в других полупроводниках, приемлемых по соображениям технологии, будут значительно более существенные изменения коэффициентов поглощения, чем те, которые представлены на рис. 3.18. Эффективные массы носителей и показатели преломления таких материалов не отличаются значительно от соответствующих значений для GaAs.

3.5.2. Устройство и характеристики ПЗС-модулятора

На рис. 3.19 представлено поперечное сечение пространственного модулятора света на основе ПЗС-структуры на GaAs. ПЗС на барьере Шоттки в этом случае является трехфазным устройством, в котором затворы, включающие одну из фаз, являются прозрачными для представляющих интерес длин волн. Излучение с длиной волны, превышающее длину волны отсечки и падающее по нормали к устройству, модулируется с помощью эффекта электропоглощения в канале ПЗС и проходит сквозь подложку. Из рис. 3.8 было видно, что пустая яма в ПЗС имеет

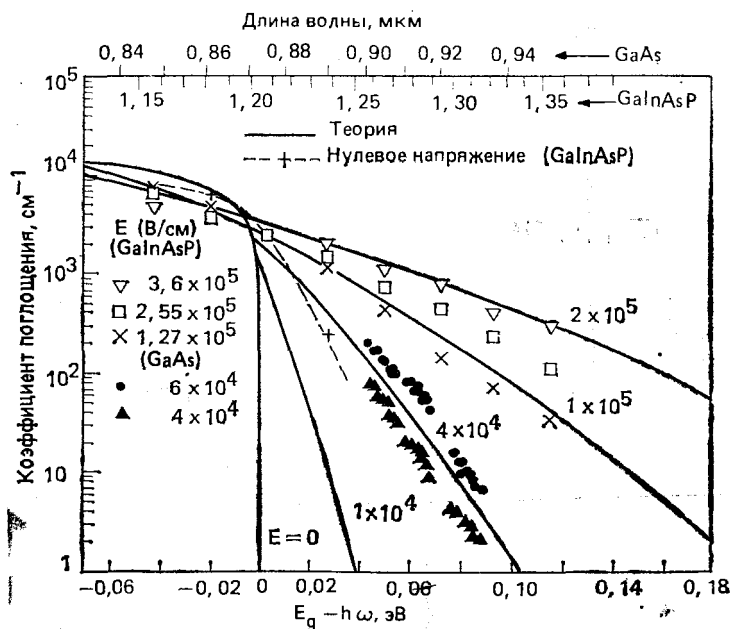


Рис. 3.18. Теоретические и экспериментальные данные по электропоглощению (эффекта Франца — Келдыша) в GaAs и GaInAsP [14].

максимальное электрическое поле, в то время как полная яма, наоборот, имеет только слабое электрическое поле под барьером Шоттки. В результате для объемного электропоглощения коэффициент оптического пропускания будет наибольшим для за-полненной ямы и наименьшим для пустой ямы.

Для вычисления модуляционных возможностей данного устройства рассмотрим упрощенную структуру на рис. 3.20, на котором показан один затвор ПЗС вместе с омическим контактом по отношению к каналу. Канал состоит из слоя n -типа с концентрацией доноров N на полубесконечной подложке. В отно-

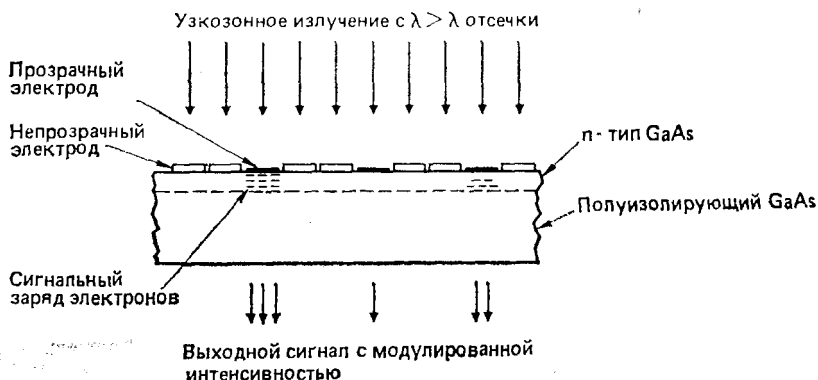


Рис. 3.19. Поперечное сечение пространственного модулятора света на ПЗС-структуре на GaAs.

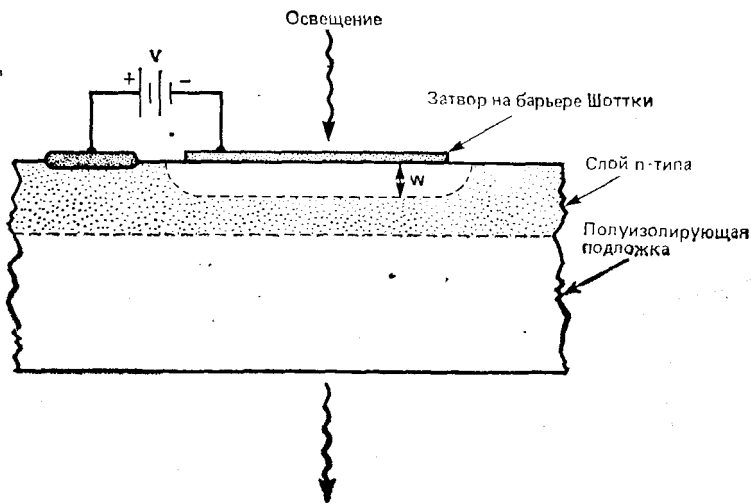


Рис. 3.20. Поперечное сечение устройства, используемого при расчетах электропоглощения,

шении максимальной модуляции эта структура не является оптимальной, однако она проста для анализа и демонстрирует основные рабочие характеристики. Когда на затвор подается отрицательный по отношению к каналу потенциал, образуется вырожденный слой глубиной ω , где ω задается выражением [29]

$$W = \left[\frac{2e}{qN} \left(V + V_{bi} - \frac{kT}{q} \right) \right]^{1/2}, \quad (3.3)$$

где V — канальное напряжение (т. е. потенциал на омическом контакте по отношению к затвору). Полагают, что канальное напряжение меньше величины V_m , необходимой для отсечки или вырождения n -слоя до границы раздела с полубесконечной подложкой. Электрическое поле $E(y)$, направленное по нормали к поверхности вырожденного слоя, задается выражением

$$E(y) = \frac{qN(y - \omega)}{\epsilon}. \quad (3.4)$$

Коэффициент пропускания оптической мощности T_0 определяется соотношением

$$T_0(\lambda, V) = \exp \left(- \int_0^{\omega} \alpha(E, \lambda) dy \right), \quad (3.5)$$

где $\alpha(E, \lambda)$ — коэффициент оптического поглощения. Заменяя переменные с помощью уравнения (3.4), можно выразить T_0 уравнением

$$T_0(\lambda, V) = \exp \left[\frac{\omega}{E(0)} \int_0^{-E(0)} \alpha(E, \lambda) dE \right]. \quad (3.6)$$

Используя данные относительно электропоглощения из рис. 3.18, можно вычислить зависимость коэффициента пропускания от канального напряжения для различных длин волн; такие данные показаны на рис. 3.21. Коэффициент пропускания был нормирован с помощью значения, полученного при нулевом напряжении канала (или условию заполненной ямы). Как было замечено ранее, коэффициент пропускания уменьшается с ростом напряжения канала. Таким же образом модуляция является наибольшей для длин волн, ближайших к длине волны отсечки 870 нм, определяемой шириной запрещенной зоны. Результаты расчетов для других уровней легирования (от $1 \cdot 10^{15}$ до $2 \cdot 10^{16}$) мало отличаются от данных на рис. 3.21.

В работе [15] сообщалось об экспериментальном подтверждении эффекта электропоглощения в ПЗС на GaAs. Модуляцию было трудно наблюдать отчасти потому, что максимальное напряжение канала составляло только 10 В, а также из-за «утечки» света по краям ямы, где электрическое поле не кон-

тролировалось зарядом. Другая проблема состоит в том, что толстая подложка (порядка 300 мкм) имела значительное поглощение, особенно на длинах волн, ближайших к краю зоны, где модуляция максимальна. Эти проблемы были решены в недавно появившемся устройстве, в котором было улучшено

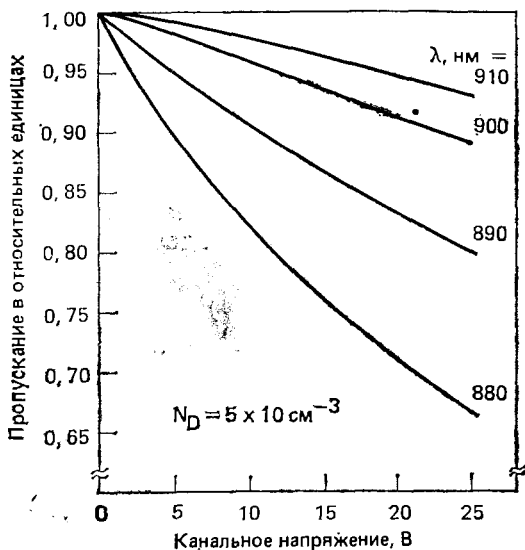


Рис. 3.21. Расчетная зависимость изменения относительного коэффициента пропускания от канального напряжения для структуры, показанной на рис. 3.20.

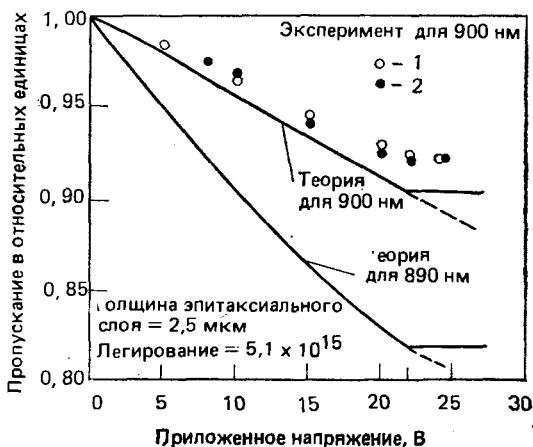


Рис. 3.22. Экспериментальные и теоретические зависимости глубины модуляции пропускания 16-каскадного ПЗС на GaAs. Вид тактовых сигналов: (1) — постоянные напряжения; (2) — обычные тактовые сигналы ПЗС.

экранирование света над затворами, повышено максимальное канальное напряжение до величин около 22 В и подложка которого была уменьшена до приблизительно 50 мкм. Модуляционные характеристики этого 16-каскадного линейного устройства показаны на рис. 3.22 для длины волны 900 нм. Измерения были проведены и для постоянных напряжений (пустые кружочки) и для обычных ПЗС тактовых сигналов (заполненные кружочки), прикладываемых к устройству. Результаты показывают несколько меньшую глубину модуляции, чем предсказано. Имеется уверенность в том, что по меньшей мере часть различия между кривыми является результатом двумерных эффектов в квантовых ямах в ПЗС, не учтенных в представленных выше одномерных расчетах. На электрические поля в каналах ПЗС сильное влияние оказывают потенциалы электродов соседних затворов, и эти накладываемые поля могут исказить форму зарядовых пакетов на краях потенциальной ямы. Следовательно, электрические поля в яме не полностью определяются зарядом в яме, соответствующим уравнению (3.4).

Хотя глубина модуляции продолжает увеличиваться с ростом напряжения канала, принципиальным ограничением этого напряжения является возникновение лавинного умножения, приводящего к заполнению ям нежелательным зарядом. Проблема лавинного заряда является особенно острой для краевых областей электродов затворов, где большие различия тактовых напряжений между смежными затворами возникают в промежутке, который с необходимостью должен быть мал, чтобы избежать проблем, связанных с неконтролируемым потенциалом под затвором. На основании имеющегося в настоящее время опыта для материалов с $N \approx 5 \cdot 10^{15}$ можно предполагать верхний предел около 25 В. При больших степенях легирования электрическое поле на поверхности $E(0)$ увеличивается для заданного канального напряжения, приводя к снижению пробойных напряжений. В общем наибольший уровень модуляции требует высоких управляющих напряжений, что в свою очередь требует использования толстых слаболегированных каналов. Практическое ограничение канальных напряжений возникает вследствие необходимости генерировать тактовые сигналы с соответственно большим размахом. По мере того, как тактовые частоты начинают превышать значение 10 МГц, становится все более трудно создать тактовую цепь ПЗС, которая могла бы генерировать на емкостной нагрузке импульсы с малым временем нарастания и спада. При этом мощность, рассеиваемая такими формирова-телями тактовых импульсов, может стать излишне большой.

Данные на рис. 3.21 показывают, что даже при лучшем стечении обстоятельств спад пропускания составит самое большее 30%. Этой величины более чем достаточно для некоторых задач обработки сигналов, таких как двумерное преобразование Фурье

[16], но, видимо, это не пригодно для задач оптических вычислений. Таким образом имеются вынужденные причины для поиска альтернативных структур, которые и позволяют усилить глубину модуляции, и снижат требования к управляющим напряжениям.

Недавно в [13] были представлены некоторые идеи относительно путей дальнейшего прогресса в данном направлении. Один важный момент, который следует отметить, состоит в том, что из-за ярко выраженной нелинейной зависимости между коэффициентом поглощения и электрическим полем, области ПЗС канала со слабым полем вносят непропорционально малый вклад в поглощение. Так, основная часть электропоглощения в структуре на рис. 3.8, *a* возникает вблизи затвора, где поле максимально. Более эффективной могла бы быть структура, в которой электрическое поле распределено однородно по большей части канала. Другим путем увеличения соотношения контраста является подход, в котором применяются многослойные структуры и управляющие электрические напряжения прикладываются к слоям параллельно, а оптическое поглощение осуществляется при последовательном прохождении слоев. Оба подхода воплощены в структуре, показанной на рис. 3.23. Основная

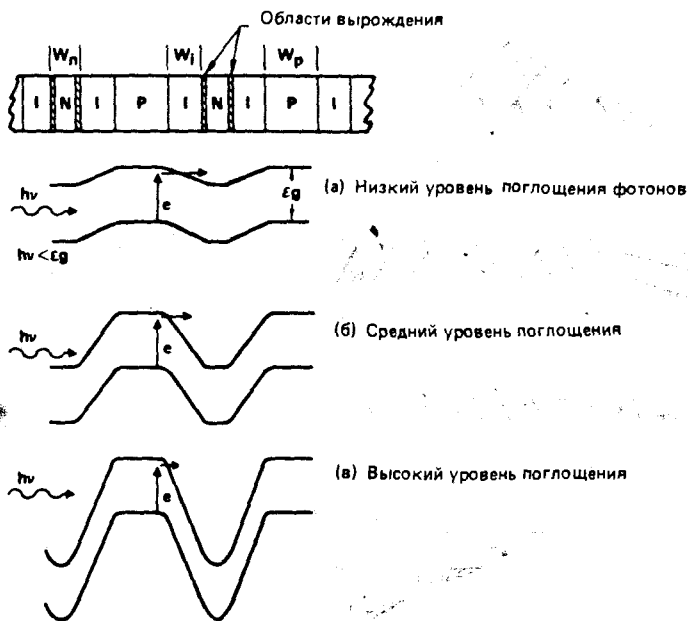


Рис. 3.23. Иллюстрация к методике увеличения контраста ПМС, основанного на электропоглощении и использующего многослойную комбинацию *p-i-n-i-p* ячеек.

ячейка состоит из последовательно легированных $p-i-n-i-p$ слоев. Слои p -типа сильно легированы и действуют как управляющие электроды или затворы, электрически разделяя каждую ячейку, в то время как в i -слое поддерживается однородное электрическое поле для получения максимального электропоглощения. Расчеты показали, что всего лишь на 4 ячейках при 20 В может быть получено затухание в 10 дБ.

Хотя этот подход привлекателен с точки зрения рабочих характеристик, он порождает значительные трудности при изготовлении устройств. Среди многих проблем изготовления этой структуры возникает необходимость создания контактов между n - и p -слоями с помощью глубоких каналов, протравленных в материале. Это требование увеличило бы, вероятно, минимальный размер элемента (50 мкм), что является нежелательным обстоятельством с точки зрения быстродействия и выхода годных устройств.

3.5.3. Эффект выбеливания

Важным ограничением возможностей этого модулятора является эффект, называемый выбеливанием и заключающийся в потере модулирующих свойств по мере того, как устройство подвергается воздействию света. Излучение, поглощаемое в ПЗС в процессе электропоглощения, создает электрон-дырочные пары. В свою очередь электроны, удерживаемые в квантовых ямах в ПЗС, уменьшают поля и, следовательно, само поглощение. В качестве грубой численной оценки возьмем типичную ячейку ПЗС, имеющую размеры порядка 30 мкм и зарядовую емкость в 10^6 электронов. Изначально незаполненная яма после поглощения 10^6 фотонов имела бы оптическое поглощение, уменьшенное с максимального до минимального значения. На длине волны, представляющей в данном случае интерес, это соответствует энергии кванта света в 0,23 пДж. Поскольку выбеливание является фундаментальным механизмом, ограничивающим возможности данного устройства, то могут быть предложены некоторые рекомендации для сведения к минимуму этой проблемы. Существенный момент состоит в том, чтобы разработать ПЗС с большой емкостью управляющего заряда. Это может быть достигнуто за счет использования ячеек большого размера и увеличения емкости затвор — канал. Последнее требует применения тонких, сильнолегированных каналов, что находится в согласии с требованиями высокого быстродействия, но противоречит требованию большой толщины поглощающей области для эффективного взаимодействия со светом.

В большей мере количественное описание эффекта можно найти на рис. 3.24. Здесь предполагается, что толщина эпитаксиального слоя, легированного до 5×10^{15} , составляет 2,1 мкм,

максимальный канальный потенциал равняется 15 В, а длина волны падающего оптического сигнала составляет 880 нм. Предполагается, что длительность потока фотонов велика по сравнению с характерным временем динамики носителей заряда, так что выбеливание «отслеживает» падающее излучение. Это предполагает использование оптических импульсов длительностью порядка пикосекунды или более. Для оптических импульсов с существенно меньшей длительностью эффекты выбеливания могли бы возникать только после того, как фотоны возбудили устройство.

На рис. 3.24 показано, как глубина модуляции изменяется по мере заполнения ям фотогенерированными носителями. Например, для изначально пустой ямы (канальное напряжение 15,0 В) заполнение ее оптическим входным сигналом на 60% емкости уменьшило бы поток фотонов до 0,85 к первоначальному значению, в то время как в отсутствие выбеливания снижение составило бы 0,76 к входному сигналу. Если яма вначале заполнена частично, тогда дополнительные фотоэлектроны могли бы переполнить яму, и оптическая модуляция больше не управлялась бы начальным канальным напряжением. Этот предел указан штриховой линией. Одно из воздействий эффекта выбеливания, следовательно, состоит в сужении динамического диапазона устройства путем ограничения диапазона размеров зарядового пакета, который может переноситься в устройстве.

3.5.4. Эффект электропоглощения в квантовых ямах

Недавно был описан второй тип электропоглощения, намного более сильный, чем эффект Франца — Келдыша [32]. Этот эффект возникает в так называемых квантовых ямах или сверхрешеточных структурах, и, как представляется, такие структуры могут быть интегрированы с ПЗС.

Квантовая яма представляет собой тонкий слой узкозонного полупроводника, размещенного между двумя слоями широкозонного полупроводника. В наиболее изученных структурах этого типа используются комбинации GaAs ($E_g = 1,42$ эВ) и $Al_xGa_{1-x}As$ ($E_g = 1,80$ эВ для $x = 0,30$). Для данного гетероперехода энергетические зоны имеют «двустороннюю» модуляцию, т. е. запрещенная зона в GaAs зависит от энергии аналогично тому, как это происходит для $Al_xGa_{1-x}As$ (рис. 3.25). Таким образом, края обеих зон — зоны проводимости и валентной зоны — образуют прямоугольные ямы, которые могут удерживать электроны и дырки. Такое удержание заряда приводит к ряду необычных новых явлений, когда ширина ямы составляет около 20 нм или менее. Представляющее здесь интерес явление касается оптического экситонного поглощения в электрических полях.

Экситоны — это связанные электрон-дырочные пары, несколько напоминающие атом водорода. Энергия фотонов, необходимая для создания экситона, составляет меньшую величину (на величину его энергии связи), чем необходимо, чтобы разорвать связанный электрон валентной зоны на свободный электрон и дырку. Их образование будет соответственно, наблюдаться как пик поглощения на длине волны немного большей, чем основной край внутрizonного поглощения. При комнатной температуре в объемном GaAs пик экситонного поглощения едва различим в основном из-за низкой энергии связи экситонов (≈ 4 мэВ), делающей их очень чувствительным к термоионизации. Однако при удержании экситонов в тонких квантовых ямах энергия их связи существенно возрастает (≈ 10 мэВ для ямы шириной в 10,0 нм), и это значительно увеличивает резонансное поглощение.

Оптическое поглощение, связанное с экситонами и другими эффектами в квантовых ямах, обычно исследуется в структурах, состоящих из большого числа ям, отделенных друг от друга слоями с большей шириной запрещенной зоны и имеющих достаточную толщину, чтобы предотвратить взаимодействие носителей в соседних ямах. На рис. 3.26 изображено устройство, включающее 60 квантовых ям, состоящих из слоев GaAs толщиной в 14,0 нм с чередующимися со слоями $\text{Al}_{0,30}\text{Ga}_{0,70}\text{As}$ толщиной 16,0 нм [11]. Эта нелегированная квантоворазмерная структура, состоящая из большого числа квантовых ям, сверху и снизу покрыта слоями *n*- и *p*-легированного AlGaAs, что позволяет прикладывать поле поперек всего блока за счет подачи

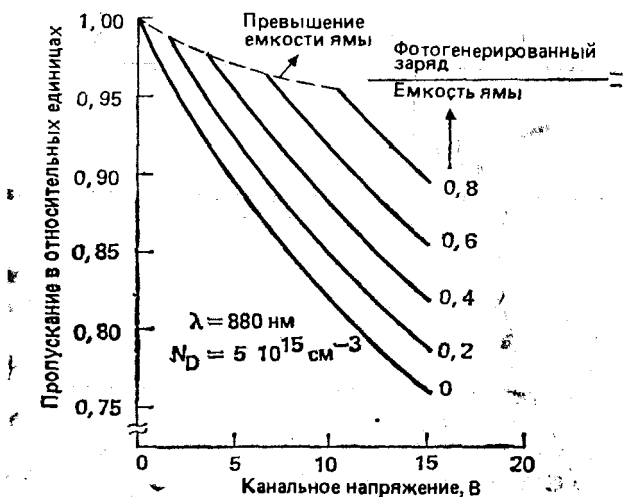


Рис. 3.24. Влияние эффекта выбеливания на динамический диапазон ПМС на основе ПЗС.

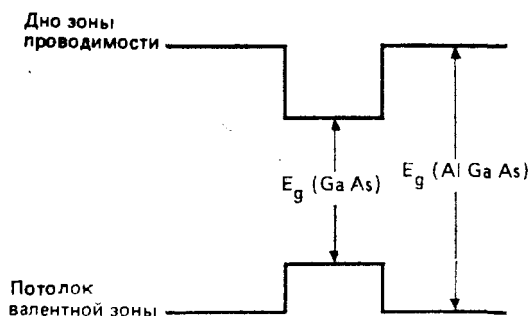


Рис. 3.25. Энергетическая зонная диаграмма квантовых ям, образованных слоем GaAs, расположенных между слоями AlGaAs.

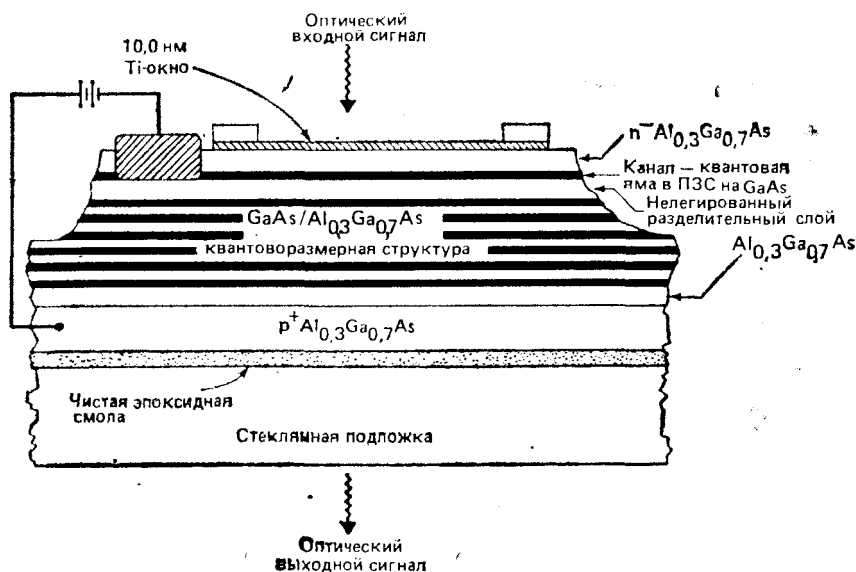


Рис. 3.26. Схема измерений коэффициентов электропоглощения квантоворазмерных структур.

на легированные слои обратной по знаку разности потенциалов. (Как объяснено ниже, дополнительные верхние слои включаются в структуру для использования в качестве канала ПЗС.) Тонкие титановые слои наносились сверху на образец, чтобы воспроизвести затвор ПЗС. Полупроводниковые слои были выращены методом молекулярной лучевой эпитаксии на подложке из GaAs. Поскольку GaAs является прозрачным на длине волны, представляющей интерес, подложку удаляли методом селективного травления и оставшаяся структура прикреплялась к стеклянной пластинке чистой эпоксидной смолой.

На рис. 3.27 показаны данные по оптическому поглощению указанной структуры [11]. Два выделяющихся пика наблюдаются на длинах волн 850 и 854 нм и связаны соответственно с экситонами, включающими тяжелые дырки и легкие дырки. В объемном GaAs может наблюдаться единственный вид экситона, но в квантовой яме повышена степень вырождения валентной зоны и возникают различные энергетические дырочные подзоны. Когда к квантоворазмерной структуре в поперечном направлении приложено напряжение, экситонные пики смещаются в сторону больших длин волны, но остаются хорошо разрешимыми. Из данных на рис. 3.27 видно, что пропускание света на длине волны экситонов, содержащих тяжелые дырки при нулевом поле, более чем удваивается при подаче 8 В. Это предполагает изменение коэффициента поглощения на 4000 см^{-1} , величину много большую, чем достижимо при эффекте электропоглощения в объеме при таких же напряженностях полей. Такие сильные эффекты также не исключены и для систем GaAs/AlGaAs. Аналогичные эффекты экситонного электропоглощения недавно наблюдались в квантоворазмерных структурах на основе InGaAs/InAlAs в диапазоне длин волн от 1100 до 1600 нм [31], используемом в волоконно-оптических линиях связи.

Надежды на объединение модулятора на квантоворазмерных структурах с ПЗС выглядят вполне обещающими. Фактически недавно была продемонстрирована успешно работавшая ПЗС, подложка которой включала квантоворазмерную структуру [12]. На рис. 3.28 показан разрез такого устройства. Уникальная особенность ПЗС — использование одиночной квантовой ямы для создания канала. Удержание носителей в тонком слое уменьшает вероятность захвата на глубокие ловушки, что представляет общую проблему для ПЗС на основе GaAs и AlGaAs. Хотя диапазон изменений канального потенциала в первом устройстве составлял менее 1 В и, следовательно, не мог быть использован для наблюдения модуляции, устройство продемонстрировало жизнеспособность концепции интеграции модуляторов на квантоворазмерных структурах и ПЗС на одной подложке.

Хотя экситонное электропоглощение создает впечатляющие перспективы, важные вопросы остаются пока без ответа. Один вопрос относится к тому факту, что в ПЗС-структуре на рис. 3.28 электрические поля не будут полностью направлены по нормали к плоскости квантоворазмерной структуры из-за различия напряжений на затворах. Если боковые компоненты поля E превышают 10^4 В/см , экситон претерпевает ионизацию полем и резонансное поглощение исчезает [7]. Таким образом, потребуется тщательно подбирать распределение полей в ПЗС, чтобы обеспечить направление электрического поля по нормали к слоям квантоворазмерной структуры, что совпадает с направлением распространения падающего света. Другой интересный

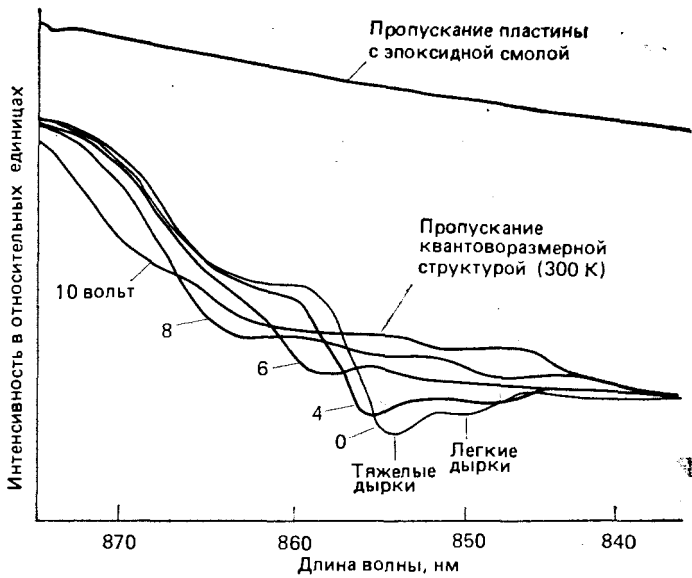


Рис. 3.27. Экспериментальная спектральная зависимость коэффициента оптического поглощения структуры, изображенной на рис. 3.26, для различных напряжений.

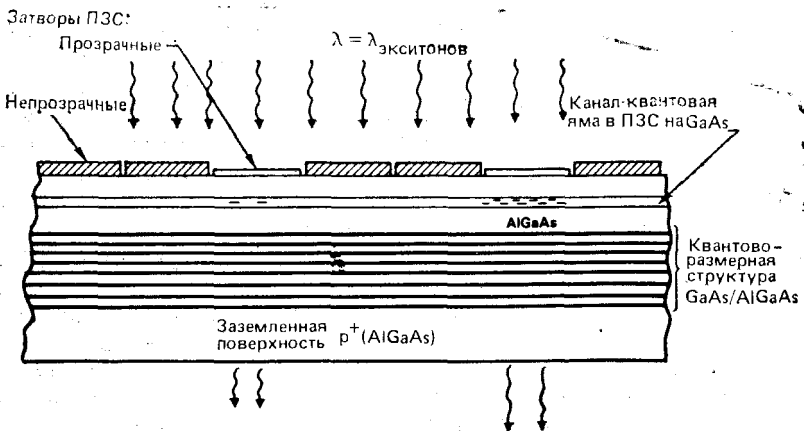


Рис. 3.28. Поперечное сечение ПМС на основе ПЗС, использующего слои, формирующие «квантовую яму», в качестве канала ПЗС и модулятор света на основе квантоворазмерной структуры.

вопрос относится к эффектам выбеливания, ожидаемым для данного устройства. Экситоны быстро ионизируются в свободные носители, которые, если они могут туннелировать или совершать надбарьерные прохождения через барьеры AlGaAs, будут уходить из квантоворазмерной структуры, что вызовет выбеливание. В описанных до настоящего времени квантоворазмерных структурах измерения фототока показывают, что свободные носители, в самом деле, могут «убегать» из квантовых ям. Вопрос о том, действительно ли данная структура может быть усовершенствована, чтобы обеспечить рекомбинацию носителей в квантоворазмерных структурах, в настоящее время является открытым и находится в стадии исследований.

3.6. Выводы

ПЗС-технология на GaAs обладает потенциальными возможностями внести значительный вклад в оптическую обработку и оптические вычисления в области матриц фотоприемников и пространственных модуляторов света. В области детекторов ПЗС предлагают значительные вспомогательные функции, такие как мультиплексирование, демultipлексирование и временное интегрирование сигналов фотоприемников. Что действительно отличается GaAs от кремния, так это продемонстрированная возможность достижения тактовых частот ПЗС, намного больших 1 ГГц, что является важным режимом для многих процессоров ЭОС. Арсенид-галлиевые ПЗС-структуры с затворами на барьерах Шоттки, такие как структуры с резистивным затвором, являются совместимыми с требованиями, предъявляемыми к широкополосным полевым транзисторам. Таким образом, возможно создание матриц фотоприемников с высоким уровнем интеграции вспомогательных цепей. Для ПМС квантоворазмерные структуры на GaAs и GaAs/AlGaAs обладают уникальными эффектами электропоглощения, не обнаруживаемыми в кремнии. Описанная здесь предварительная работа показала, что адресуемые ПЗС-квантоворазмерные структуры являются многообещающими для получения быстродействующих двумерных ПМС, удовлетворяющих требованиям к значениям экстинкции и, возможно, к уровню эффектов выбеливания. Однако в этом случае необходимо дополнительное понимание физики этого устройства.

За исключением квантоворазмерных структур, были созданы основные принципы и варианты этих устройств в уменьшенных масштабах. Основная задача остается в усовершенствовании конструкций устройств, повышении качества материалов и создании больших одно- и двумерных матриц. К сожалению, материалы и технология изготовления GaAs и AlGaAs находятся на низком уровне развития по сравнению с кремнием и

это является главным препятствием для технологии описанных выше устройств. Недавний всплеск интереса к интегральным схемам на GaAs дает надежды на то, что однажды появятся матрицы детекторов со сверхвысокой степенью интеграции и ПМС, созданные на основе ПЗС-структур.

В доказательство этого следует заметить, что недавно были продемонстрированы ПЗС-пространственные модуляторы света, использующие квантоворазмерные GaAs/AlGaAs и GaAs/InGaAs структуры [2, 12, 21].

МНОГОЗНАЧНАЯ И ПОРОГОВАЯ ЛОГИКА

Глава 4

МНОГОЗНАЧНАЯ ЛОГИКА В ОПТИЧЕСКИХ ВЫЧИСЛЕНИЯХ

*М. Коннер, Д. Эйхманн, Отделение электротехники,
Городской колледж Университета г. Нью-Йорка,
Нью-Йорк, шт. Нью-Йорк*

4.1. Введение

Здравый смысл подсказывает, что очень трудно создать оптическую аналоговую вычислительную систему с динамическим диапазоном, намного большим чем 30 дБ. Следовательно, поиск ведется в области дискретных систем, где сигналы задаются числами, а количество разрядов числа определяет динамический диапазон системы. Таким образом, динамический диапазон системы полностью зависит от воли разработчика. Наиболее широко распространенным основанием таких численных представлений, несомненно, является цифра два, приводящая к разработке двоичной логики.

Двоичная логика, по крайней мере в области электроники, имеет длинную историю и возникла из двоичных свойств реле. Реле — это элемент, который либо открыт, либо закрыт. Наиболее широко используемые электронные устройства, такие как вакуумные трубки, транзисторы и т. д., также имеют два четко определенных состояния: выключенное и насыщенное. Логика на основе насыщающихся элементов работала достаточно хорошо. Однако вскоре было обнаружено, что использование этих элементов ограничивает быстродействие систем. Результатом данного обстоятельства явилось возникновение логических схем с эмиттерными связями. И как только появилось желание отказаться от насыщающихся логических элементов с характерными для них защищенностью от шума и простотой конструкции ради достижения более высокого быстродействия, то, естественно, возник вопрос: «А почему бы заодно не отделаться и от двоичного кода?». Ответ в данном случае определялся двумя соображениями, во-первых, тем, что при заданном динамическом диапазоне двоичная логика наименее восприимчива к шуму, и, во-вторых, приверженностью к традициям. Ведь это очень привычно — конструировать схемы на основе двоичных элементов. Однако с появлением оптических вычислений с этой традицией приходится порывать, поскольку теперь логические операции строятся на иной основе, нежели электрический ток и напряжение. Вместо этого для проведения вычислений изу-

чают возможности новых физических явлений. Когда обсуждаются оптические дискретные системы, то многозначная логика, чьи логические компоненты представляют собой многочисленные значения входных и выходных элементов вместо только двух, должна рассматриваться как реальная альтернатива обычной двоичной логике. Цель данной главы состоит в том, чтобы начать диалог об эффективности многозначных оптических логических систем.

При разработке любой логической схемы первоочередной задачей является выбор логических элементов, которые следует использовать. Так, например, может быть использован ряд канонических двоичных множеств логических элементов. Чтобы сделать наше обсуждение условий вхождения логического элемента в каноническую систему более живым, в разд. 4.2 дано краткое описание проблемы полноты двоичной логики. Этот вопрос, обобщенный до представлений о полноте многозначной логики, является решающим при определении, когда группа оптических явлений может рассматриваться как часть канонического множества оптических логических элементов. В разд. 4.3 описан специфический пример многозначной логической системы, обладающей слабой полнотой, — системы счисления в остаточных классах (ССОК). Еще совсем недавно алгебра ССОК рассматривалась применительно к арифметическим вычислениям в остаточных классах. По вопросу оптической реализации различных операций в ССОК имеется большое число публикаций, обзор которых сделан в разд. 4.4. Оптические элементы могут образовывать стандартные блоки оптической многозначной логической схемы. В заключительном, в значительной мере техническом разделе описаны некоторые из необходимых тестов, служащих для установления принадлежности многозначной логической функции каноническому множеству. В этом случае такие многозначные логические функции и их оптическая реализация могли бы послужить новыми элементами оптических многозначных логических схем.

4.2. Проблема полноты в двоичной логике

Перед тем как приступить к обсуждению общих вопросов многозначной логики, рассмотрим несколько более простой двоичный случай. Желая ознакомиться с этими вопросами более глубоко могут обратиться к работам [1, 2]. В двоичной логической системе имеются два определенных логических уровня. Эти уровни обычно обозначаются как 0 и 1. В общем случае комбинаторная логическая система будет иметь n входных сигналов, обозначаемых как вектор $x = (x_1, x_2, \dots, x_n)$ и m выходных сигналов, обозначаемых как вектор $y = (y_1, y_2, \dots, y_m)$. Элементы каждого из векторов могут принимать значения 0

или 1. Корректные схемы комбинаторной логики не требуют обратной связи выхода со входом, или любых соединений между выходными элементами [1—3]. Релейно-контактная схема f просто преобразует входной сигнал x в выходной сигнал y , т. е. $y = f(x)$. Упростим обсуждение и обратимся к случаю большого числа входных сигналов и одного выходного. Задание таблицы истинности и реализацию желаемых выходных значений на основе заданных входных сигналов обычно проводят на основе практики. Аппаратное обеспечение реализации заданной таблицы истинности вытекает из комбинации определенных стандартных элементов. Простые релейно-контактные компоненты, из которых строят функции произвольной сложности, называют *элементарными логическими функциями* (ЭЛФ). Здесь уместен вопрос о том, когда множество ЭЛФ является полным, т. е. может ли произвольная логическая функция быть представлена корректной комбинацией элементов из множества ЭЛФ?

Для того чтобы установить условия логической полноты, требуется определить пять свойств ЭЛФ: монотонность, линейность, самодвойственность, сохранение 0 и 1. Для того чтобы дать определенное монотонной ЭЛФ, требуется определить понятие упорядоченности. Рассмотрим два двоичных n -мерных вектора $\mathbf{a} = (a_1, a_2, \dots, a_n)$ и $\mathbf{b} = (b_1, b_2, \dots, b_n)$. Теперь $\mathbf{a} \leq \mathbf{b}$, если каждый элемент \mathbf{a} меньше, либо равен в обычном арифметическом смысле, соответствующего элемента \mathbf{b} . ЭЛФ f монотонна, если при $\mathbf{a} < \mathbf{b}$ и $f(\mathbf{a}) = 1$ справедливо $f(\mathbf{b}) = 1$. ЭЛФ f линейна, если она может быть записана в виде двоичного линейного полинома

$$f(x_1, x_2, \dots, x_n) = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \pmod{2}, \quad (4.1)$$

где a_i — константы, принимающие значение либо 0, либо 1. ЭЛФ f является самодвойственной, если в результате подачи на вход дополнения ее входного сигнала на выходе получаем дополнение ее выходного сигнала для всех возможных входных сигналов, т. е.

$$f(x_1, x_2, \dots, x_n) = \bar{f}(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n).$$

Дополнением 0(1) служит 1(0). ЭЛФ f сохраняет 0(1), если, подавая на все входы 0(1), на выходе получим 0(1). Примером монотонной функции может служить логическое И, а логическое ИСКЛЮЧАЮЩЕЕ ИЛИ не является таковой. Постоянная функция $f=1$ и ИСКЛЮЧАЮЩЕЕ ИЛИ линейны, а ЭЛФ ИЛИ-НЕ не обладает линейностью. ЭЛФ НЕ является самодвойственной, но И таковым не является. Однако И сохраняет и 0, и 1, в то время как ИЛИ-НЕ не сохраняет 0, а ИСКЛЮЧАЮЩЕЕ ИЛИ не сохраняет 1.

В качестве примера проверим двоичную переключающую функцию на выполнение пяти указанных выше условий. Рас-

Таблица 4.1. Таблица истинности для двоичной переключающей функции из примера 4.1

	x_1	x_2	x_3	f_1
a_1	0	0	0	1
a_2	0	0	1	0
a_3	0	1	0	0
a_4	0	1	1	1
a_5	1	0	0	0
a_6	1	0	1	1
a_7	1	1	0	1
a_8	1	1	1	0

смотрим приведенную ниже функцию трех логических переменных, потенциально являющуюся ЭЛФ.

Пример 4.1

Рассмотрим табл. 4.1. Проверка каждого из рядов показывает, что f_1 не монотонна. Проводя непосредственные алгебраические действия, обнаружим, что f_1 может быть выражено как

$$f_1 = 1 + x_1 + x_2 + x_3 \pmod{2} \quad (4.2)$$

и, следовательно, f_1 является и линейной, и самодвойственной. Проверка строки a_1 показывает, что «не сохраняется нуль», в то время как проверка строки a_8 показывает, что f_1 также «не сохраняет 1».

После того как ЭЛФ определены, можно, опираясь на работу [4], ответить на вопрос, касающийся их полноты. Множество ЭЛФ является полным, если оно содержит хотя бы одну элементарную функцию из следующих пяти классов:

- ЭЛФ, не сохраняющая 0;
- ЭЛФ, не сохраняющая 1;
- ЭЛФ, не являющаяся самодвойственной;
- ЭЛФ, не являющаяся линейной;
- ЭЛФ, не являющаяся монотонной.

Доказательство достаточности этих условий требует определенных математических выкладок, а необходимость может быть установлена непосредственно. Заметим, что если все ЭЛФ, например, сохраняют 0, то очевидно, что функции, взятые от функций, сохраняющих 0, также будут сохранять 0. Однако если из заданного множества требуется получить такую логическую функцию l , что $l(0, 0, \dots, 0) = 1$, то эта функция не может быть синтезирована. Например, элементарные логические функции И-НЕ и ИЛИ-НЕ обе являются полными.

Если отказаться от требования, что можно получить постоянные логические функции $f=0$ и $f=1$, используя ЭЛФ, тогда условие получения всех возможных логических функций из множества ЭЛФ относят к критерию слабой полноты системы логических функций. Слабую полноту можно было бы рассматривать как более практичное соображение, потому что разработчик логики обычно осуществляет контроль по постоянным значениям входного сигнала. Множество ЭЛФ является слабо полным тогда и только тогда, когда оно содержит ЭЛФ, обладающие двумя свойствами: одну ЭЛФ, не являющуюся монотонной, и одну — не являющуюся линейной.

Очевидно, что любое множество, являющееся полным, одновременно является и слабо полным. ИСКЛЮЧАЮЩЕЕ ИЛИ никогда не является ни полной, ни слабо полной. ЭЛФ ИСКЛЮЧАЮЩЕЕ ИЛИ и И являются слабо полными. Обе функции сохраняют 0. Однако И является нелинейной, а ИСКЛЮЧАЮЩЕЕ ИЛИ — немонотонной. Отдельные ЭЛФ, такие как И-НЕ и ИЛИ-НЕ, удовлетворяющие требованиям слабой полноты, называются универсальными ЭЛФ. На этом завершается короткое обсуждение возможных канонических логических элементов для систем двоичной логики. В следующем разделе будет рассмотрена одна из возможных многозначных систем, обладающих слабой полнотой — модульные вычисления в ССОК. Такая система может быть реализована в оптике различными способами. Некоторые из этих способов будут описаны в последующих разделах.

4.3. Многозначные логические схемы, основанные на системе счисления в остаточных классах (ССОК)

Обсуждение многозначных логических элементов начнем с рассмотрения обладающего слабой полнотой множества элементов для выполнения операций сложения и умножения по модулю p , являющемуся простым целым числом. Операции сложения и умножения по модулю p образуют поле тогда и только тогда, когда p является простым числом: существует только p различных элементов, $\{0, 1, 2, \dots, p-1\}$. Эта часть обсуждения заимствована из работы [5]. Рассмотрим произвольную функцию f , имеющую значение $f(0) = c_0, f(1) = c_1, \dots, f(p-1) = c_{p-1}$. Всегда может быть составлен полином вида

$$f(x) = \sum_{i=0}^{p-1} a_i x^i \pmod{p}. \quad (4.3)$$

Чтобы убедиться в этом, рассмотрим следующую систему уравнений:

Таблица 4.2. Обобщенная спецификация таблицы истинности для многозначной переключающей функции двух переменных $f(x, y)$

y/x	0	1 ...	$p-1$
0	c_{00}	c_{01}	$c_{0(p-1)}$
1	c_{10}	c_{11}	$c_{1(p-1)}$
.	.	.	.
.	.	.	.
$p-1$	$c_{(p-1)0}$	$c_{(p-1)1}$	$c_{(p-1)(p-1)}$

$$c_0 = a_0$$

$$c_1 = a_0 + a_1 + a_2 + \dots + a_{p-1}$$

$$c_2 = a_0 + 2a_1 + 2^2a_2 + \dots + 2^{p-1}a_{p-1}$$

.

.

.

$$c_{p-1} = a_0 + (p-1)a_1 + (p-1)^2a_2 + \dots + (p-1)^{p-1}a_{p-1}. \quad (4.4)$$

В матричной форме эта система выглядит как: $\mathbf{c} = V\mathbf{a}$, где \mathbf{c} и \mathbf{a} вектор-столбцы, $\mathbf{c} = (c_0, c_1, \dots, c_{p-1})^T$ и $\mathbf{a} = (a_0, a_1, \dots, a_{p-1})^T$, где T обозначает транспонирование, а V — матрица Вандермонда [6].

$$V = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 2^2 & \dots & 2^{p-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & p-1 & (p-1)^2 & \dots & (p-1)^{p-1} \end{vmatrix}. \quad (4.5)$$

Матрица V является обратимой, потому что вычисления детерминанта (см. [5]) дают величину

$$\det V = 1!2!3! \dots (p-1)! \neq 0 \pmod{p}, \quad (4.6)$$

являющуюся ненулевой, если p — простое число. Таким образом, видим, что любая переключающая многозначная функция одной переменной в ССОК может быть представлена как полином порядка $(p-1)$ по модулю p .

Этот результат может быть распространен на многозначную переключающую функцию двух переменных $f(x, y)$. Для простого p каждая из p^{p^2} функций двух переменных может быть

определена в табл. 4.2, являющейся эквивалентом полинома вида

$$f(x, y) = f_0(x) + f_1(x)y + f_1(x)y^2 + \dots + f_{p-1}(x)y^{p-1} \pmod{p},$$

где

$$f_i(x) = a_{0i} + a_{1i}x + a_{2i}x^2 + \dots + a_{(p-1)i}x^{p-1} \pmod{p}. \quad (4.76)$$

Пример 4.2

Рассмотрим трехуровневую переключающую функцию двух переменных $f(x, y)$ в ССОК, определенную в табл. 4.3. Данная

Таблица 4.3. Трехуровневая переключающая ССОК-функция двух переменных

y/x	0	1	2
0	1	1	0
1	2	0	2
2	2	0	2

функция может быть представлена как полином двух переменных, где под операциями понимаются операции по mod 3

$$f(x, y) = 1 + 2x + x^2 + y^2 + x^2y^2 \pmod{3}.$$

Пример 4.3

Рассмотрим более сложную трехуровневую переключающую функцию $f(x, y)$, определенную в табл. 4.4. Ниже представлен

Таблица 4.4. Вариант переключающей ССОК-функции двух переменных

y/x	0	1	2
0	1	2	0
1	2	2	1
2	0	0	1

непосредственный способ получения переключающего полинома. В общем виде полином будет иметь вид

$$f(x, y) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy + a_6x^2y + a_7xy^2 + a_8x^2y^2.$$

Вычислим коэффициенты a :

$$\begin{array}{l} f(0, 0) = 1 \quad a_0 = 1 \\ f(0, 1) = 2 \quad \text{и} \quad 1 + a_2 + a_4 = 2 \\ f(0, 2) = 0 \quad 1 + 2a_2 + a_4 = 0 \\ \hline 2 + \quad 2a_4 = 2 \quad a_4 = 0 \\ a_2 = 1 \end{array}$$

$$\begin{array}{l} f(1, 0) = 2 \quad 1 + a_1 + a_3 = 2 \\ f(2, 0) = 0 \quad 1 + 2a_1 + a_3 = 0 \\ \hline 2 + \quad 2a_3 = 2 \quad a_3 = 0 \\ a_1 = 1 \end{array}$$

$$\begin{array}{l} 1 + x + y + a_5xy + a_6x^2y + a_7xy^2 + a_8x^2y^2 \\ f(2, 2) = 1 \quad 2 + a_5 + 2a_6 + 2a_7 + a_8 = 1 \\ f(2, 1) = 1 \quad 1 + 2a_5 + a_6 + 2a_7 + a_8 = 1 \end{array}$$

$$\begin{array}{l} + \\ f(1, 2) = 0 \quad 1 + 2a_5 + 2a_6 + a_7 + a_8 = 0 \end{array}$$

$$\begin{array}{l} + \\ f(1, 1) = 2 \quad 0 + a_5 + a_6 + a_7 + a_8 = 2 \\ \hline 1 \quad a_8 = 1 \end{array}$$

$$a_8 = 0$$

$$1 \quad 2a_7 = 2$$

$$2a_7 = 2$$

$$2 + a_5 = 1$$

$$a_5 = 2$$

$$2 + 2 + 2a_6 + 1 + 0 = 1$$

$$2a_6 + 2 = 1$$

$$2a_6 = 2$$

$$a_6 = 1$$

Следовательно, в результате получается полином

$$f(x, y) = 1 + x + y + 2xy + x^2y + 2xy^2 \pmod{3}.$$

Один из способов убедиться в том, что эта процедура всегда работает,— это рассмотреть окончательный вид табл. 4.5. Рядом с входными сигналами функции приведены произведения четырех термов, составленных так, чтобы произведение равнялось нулю, если значения x и y не удовлетворяют соответствующему входному сигналу. Чтобы получить значение функции, каждый член вычисляется при соответствующем входном сигнале и умножается на соответствующую константу. Теперь для получения окончательного вида полинома можно было бы перемножить

Таблица 4.5. Второй способ представления трюичных переключющих функций из табл. 4.4

x	y	f		
0	0	1	$(x+1)(x+2)(y+1)(y+2)$	1
0	1	2	$(x+1)(x+2)(y)(y+1)$	2
0	2			0
1	0	2	$x(x+1)(y+1)(y+2)$	2
1	1	2	$x(x+1)(y)(y+1)$	2
1	2	0		0
2	0	0		0
2	1	1	$(x)(x+2)(y)(y+1)$	1
2	2	1	$(x)(x+2)(y)(y+2)$	1

между собой все члены в выражениях и сложить по модулю 3.

Отметим, что методы преобразования многозначной таблицы истинности в полиномы не приводят к минимальным выражениям для аппаратной реализации. Рассмотрим пример 4.2, в котором был получен полином

$$f(x, y) = 1 + 2x + x^2 + y^2 + x^2y^2 \pmod{3}. \quad (4.8a)$$

Если бы двумя типами устройств, которые необходимо использовать, являлись 2-входовые сумматоры и умножители, тогда достаточно простая реализация могла бы выглядеть так, как показано на рис. 4.1. Однако не сложно обнаружить, что $f(x, y)$ может быть преобразована к виду

$$\begin{aligned} f(x, y) &= 1 + 2x + x^2 + y^2 + x^2y^2 \\ &= (x+1)^2 + y^2(x^2+1) \pmod{3}. \end{aligned} \quad (4.8b)$$

Если реализовать непосредственно это выражение, то получим схему из трех сумматоров и четырех умножителей. Чтобы при неизменном числе сумматоров уменьшить число умножителей до 3, необходимо отметить, что

$$x^2 + 1 = (x+1)^2 = 2x = (x+1)^2 + x \pmod{3}.$$

Схема для последнего выражения показана на рис. 4.2. К сожалению, не существует общих алгоритмов работы с полиномиальными уравнениями, позволяющих достичь минимальной степени сложности схем; в данном случае следует полагаться на способности и интуицию разработчика.

Вторым способом построения необходимой многозначной переключющей функции $f(x, y)$ является использование интерполяционных полиномов $J_i(x)$. Эти полиномы определяются как

$$J_i(x) = \begin{cases} 1, & \text{если } x=i, \\ 0 & \text{в противном случае.} \end{cases} \quad (4.9a)$$

С помощью теоремы Ферма [7] эта функция может быть выражена как $J_0(x) = (p-1)x^{p-1} + 1$ и

$$J_i(x) = (p-1)[x^{p-1} + ix^{p-2} + i^2x^{p-3} + \dots + i^{p-2}x], \quad (4.96)$$

где $1 \leq i \leq p-1$. Для переключающей функции n переменных, где значение функции при $\mathbf{x}=\mathbf{a}$ составляет $f(\mathbf{a})=K_i$. Окончательное выражение имеет вид

$$f(\mathbf{x}) = K_0J_0(\mathbf{x}) + K_1J_1(\mathbf{x}) + \dots + K_{p-1}J_{p-1}(\mathbf{x}), \quad (4.10a)$$

где

$$J_a(\mathbf{x}) = J_{a_1}(x_1) J_{a_2}(x_2) \dots J_{a_{p-1}}(x_{p-1}). \quad (4.10b)$$

Пример 4.4

Элементарными функциями $J_i(x)$ по модулю 3 являются

$$J_0(x) = 2x^2 + 1,$$

$$J_1(x) = 2(x^2 + x) = 2x^2 + 2x,$$

$$J_2(x) = 2(x^2 + 2x) = 2x^2 + 4x.$$

Если умножение производится по модулю p , являющемуся простым числом, то в ССОК модульное умножение может быть сведено к операции сложения. Процедура основывается на представлении в степенном виде. С помощью теоремы Ферма [7] можно показать, что существует такое именуемое генератором число b , что, будучи возведенным в степень $p-1$, дает 1 по mod p , т. е.

$$b^{p-1} = 1 \pmod{p}. \quad (4.11)$$

Для этих чисел составлены таблицы (см. [23]). Возводя генератор в соответствующие степени, все числа могут быть представлены в виде выражений по модулю p . При использовании в ССОК этих экспоненциальных представлений чисел умножение может быть заменено сложением показателей степеней. Можно показать, что это сложение имеет модуль $p-1$.

Пример 4.5

Умножим 3 на 4 по модулю 5, используя метод сложения показателей степеней. Генератором по модулю 5 является 2. Таким образом, используя этот генератор, получаем

$$2^0 = 1, \quad 2^1 = 2, \quad 2^2 = 4, \quad 2^3 = 3, \quad 2^4 = 1.$$

Заметим, что, во-первых, следуя теореме Ферма, «наибольшее» значение показателя степени составляет $p-1=4$, а, во-вторых, нуль не может быть получен, и, следовательно, эта процедура должна быть учтена отдельно. Таким образом, для того чтобы умножить 3 на 4, складывают соответствующие им значения показателей степени, а именно $3+2=1 \pmod{4}$. Данный пока-

затель степени, являющийся эквивалентом логарифма, соответствует значению 2.

Процедура генерирования значения показателя степени для заданного генератора и простых модулей называется прямым K -преобразованием. В свою очередь генерирование символов из значений показателей степени называется обратным K -преобразованием. И то и другое преобразование эквивалентно операции перестановки. Если операции перестановки допустимы, то необходимо только модульное сложение.

Закончим краткий экскурс в многозначную логическую алгебру ССОК и перейдем к рассмотрению оптической реализации операций ССОК.

4.4. Оптические многозначные логические ССОК-процессоры

В данном разделе будет рассмотрен ряд различных архитектур оптической многозначной логической обработки сигналов. Как отмечалось в предыдущих разделах, существует целый ряд полных систем множеств многозначной логики. Здесь будет дан обзор только некоторых из возможных архитектур оптической обработки в ССОК, а именно — являющихся кандидатами для многозначной логической обработки. По данному вопросу име-

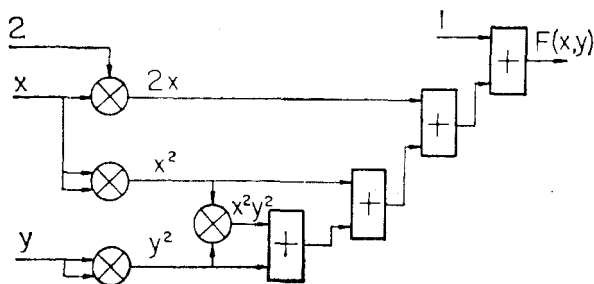


Рис. 4.1. Простая схема реализации переключающей функции из табл. 4.4 на основе сумматоров и умножителей по mod 3.

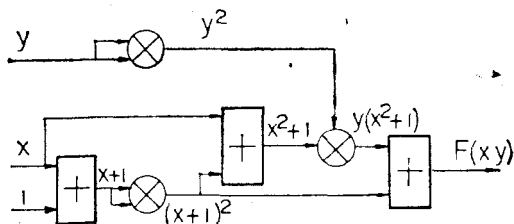


Рис. 4.2. Упрощенная схема реализации переключающей функции из табл. 4.4.

ется целый ряд интересных публикаций [8—30]. После рассмотрения возможных методик будет описано несколько новых оптических ССОК процессоров и их приложение для многозначной логической обработки.

Один из самых ранних электрооптических процессоров был сконструирован Д. Н. Лемером [8]. В данном процессоре для открывания и закрывания переключателей, выполнявших функции механического «сита», использовались механические шестеренки, по одной на каждое простое число, представлявшее отдельную периодическую операцию. Световой пучок с помощью фотоэлектрического детектора регистрировал состояние фильтра. Это устройство способно решать алгебраические уравнения с простыми модулями и использовано для разложения натуральных чисел на простые множители. Хорошо известно, что любое натуральное число может быть представлено в виде произведения простых чисел. В арифметических ССОК-процессорах каждое из этих простых чисел в свою очередь представляет собой независимый канал арифметики остаточных классов. Арифметические операции, такие как сложение, вычитание, умножение и деление, выполняются единственным образом над двумя натуральными числами. При этом используется представление числа в виде остатков такой длины, чтобы результат находился внутри диапазона, определяемого произведением абсолютных величин произведения первичных остатков. Тот факт, что арифметические операции в ССОК не требуют переноса между рядами, т. е. каждый канал ССОК работает независимо, делает эти процессоры особенно привлекательными с точки зрения оптических вычислений. Надо сказать, что процессоры, работающие в ССОК, с недавних пор вызывают все больший интерес, и это объясняется одновременно двумя фактами — и тем, что они не требуют переноса, и тем, что арифметические операции выполняются с использованием независимых каналов [9—21].

Для оптического многозначного логического ССОК-процессора необходимы две арифметические операции над остатками — сложение и умножение остатков по модулю, являющемуся простым числом. Операция умножения остатков может рассматриваться как повторяемое сложение остатков. В качестве альтернативы операция умножения остатков по простому модулю p может быть сведена к сложению остатков по модулю $p-1$ [7]. Данный процесс близок к комбинированному процессу возведения в степень, сложения и логарифмирования, используемым для умножения действительных чисел с помощью процесса сложения. Необходимыми преобразованиями, обозначенными как K -преобразования, являются операции перестановки, которые легко могут быть выполнены оптически. Следовательно, основной акцент в оптических операциях ССОК будет делаться на оптическом сложении в ССОК.

Существуют два основных подхода, которые можно использовать для выполнения оптическими методами сложения в ССОК: по своей сути это оптические аналоговые и оптические цифровые методы. Аналоговые оптическое кодирование и операции с символами остатков чисел используют периодические величины, такие как пространственная или временная фаза оптического поля. В оптическом цифровом кодировании любой из его видов, включая пространственное (обычно являющееся оптической модуляцией координаты), угловое, временное, частотное и амплитудное кодирование, может быть использован для представления необходимых символов остатков чисел. Эти символы в свою очередь включаются в работу с помощью специально сконструированных оптических систем. Эти системы обладают либо пространственной, либо временной периодичностью, которая помогает свести арифметические операции к периодическим величинам. Тогда арифметические операции, например для оптической координатной модуляции, эквивалентны сдвигу световых пятен с одного места на другое. Эта операция сдвига в свою очередь может быть визуализирована двумя способами. В первом способе подготавливают определенные наборы смещенных друг относительно друга элементов, также называемых картами. В зависимости от требуемого режима световое пятно перемещают на определенную карту. Эти карты являются эквивалентными оптическим поисковым таблицам. Во втором режиме работы перемещают сами карты.

Наиболее простым способом оптической кодировки различных модулированных¹⁾ чисел ССОК является использование аналоговых периодических явлений, таких как временная или пространственная фазовая модуляция. Сложение двух чисел ССОК тогда равняется сложению фаз, которые сами являются величиной с модулем 2π . Эти оптические методики были описаны в работах [11—13]. В схеме, предложенной Коллинзом, операторы арифметического сложения в ССОК выполнены на дискретных состояниях поляризации оптического луча. Модуляция поляризации может быть также выполнена с помощью двумерного пространственного модулятора света (ПМС). Здесь (рис. 4.3) входной пучок с известной заданной поляризацией отражается от поверхности ПМС. Числа в ССОК, которые необходимо складывать, вводятся сзади в ПМС в виде модулированной интенсивности. Эти интенсивности модулируют внешнюю поверхность ПМС, вызывая дискретные изменения поляризации отраженного от передней поверхности пучка. При повторном отражении с помощью зеркала от передней поверхности ПМС поляризация может быть успешно изменена за счет

¹⁾ Имеется в виду представление числа по различным модулям. —
Прим. ред.

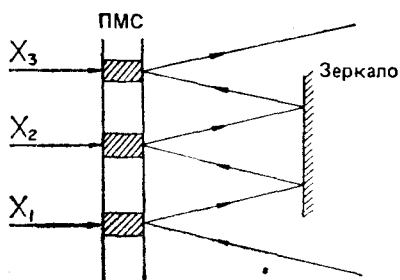


Рис. 4.3. Модульное кодирование операции сложения в ССОК с помощью модуляции поляризации света. Модуляция поляризации осуществляется с помощью двумерного пространственного модулятора света [11].

сопутствующего добавления различных фазовых углов. Это кодирование является полностью аналоговым и поэтому чувствительным к воздействию шума. Быстродействие этого процессора ограничено тактовой частотой ПМС. Хотя в настоящее время эти частоты довольно низки, тот факт, что для параллельной обработки можно использовать двумерную поверхность, может привести к значительному повышению производительности.

Вследствие своей относительной помехозащищенности оптическая модуляция по координате является более предпочтительным методом оптического кодирования различных символов чисел в ССОК. Пространственное расположение символов в этом случае определяет «значение» цифры остатка. Имеется ряд известных методик, использующих этот метод кодирования операций в ССОК. Будут описаны несколько методов; в некоторых используются фиксированные карты, или перекодировочные таблицы, в то время как в других используют динамически изменяющиеся карты. Часть оптических методик реализации перекодировочных таблиц относится к калейдоскопической оптике [14], голографическим перекодировочным таблицам [15] и различным устройствам переключения луча [16]. Часть динамических методик относится к корреляционным способам [17] и различным электрооптическим волноводным переключателям [14, 18].

Простейший способ изменять положение заключается в использовании оптических дефлекторов для отклонения светового пятна. В этих методиках для изменения положения светового пятна используется произвольный набор зеркал, решеток и призм (рис. 4.4). Эти устройства могут быть уменьшены до размера интегральной схемы. В дополнение для переноса светового пучка в пространстве в качестве световодов могут быть использованы волоконно-оптические элементы. В калейдоскопическом оптическом устройстве зеркальный прямоугольный волновод используется для изменения направления падающего света в пространстве (рис. 4.5). При правильном выборе конструкции могут быть реализованы различные распределения пространственной оптической карты. Этот тип оптики имеет спо-

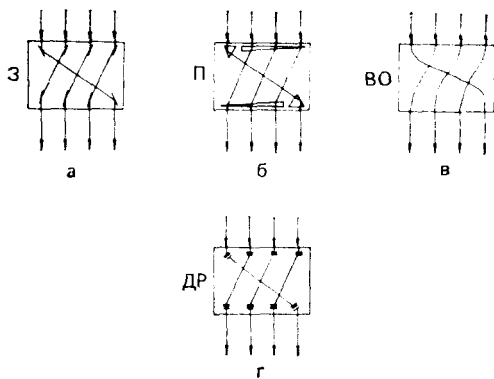


Рис. 4.4. Методики кодировки пассивных пространственных карт [15]: а — с помощью зеркал; б — с помощью призм; в — с помощью волоконной оптики; г — с помощью дифракционных решеток.

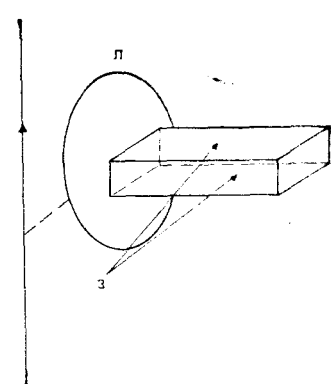


Рис. 4.5. Калейдоскопическое устройство для пространственного периодического кодирования светового пятна [13].

способность уменьшать пространственные апертуры, переводя непериодический диапазон чисел в подходящий периодический диапазон пространственных направлений. Голографические перекодировочные таблицы осуществляют перевод входного числа в ССОК, обычно представляемого двоичной записью десятичного числа, с набором предварительно изготовленных оптических согласующих фильтров-масок (рис. 4.6). Маски представляют все возможные комбинации переменных, которые могут присутствовать во входном сигнале. Требуемая маска выбирается посредством установления угла опорного пучка. Система обычно сконструирована таким образом, что только выбранный выходной сигнал несет необходимую корреляцию, т. е. для заданного входного сигнала все остальные комбинации масок иск-

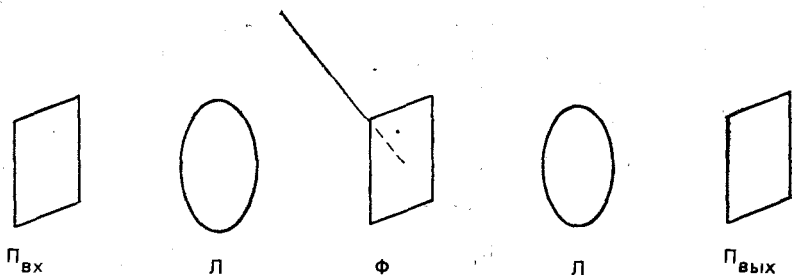


Рис. 4.6. Перекодировочная таблица с голографическим управляемым фильтром, предназначенная для оптического кодирования различных операций ССОК [14].

$\Pi_{вх}$ — входная плоскость; Л — линза; Ф — фильтр; $\Pi_{вых}$ — выходная плоскость.

лучаются. Имеются две различные системы: оптическая операция ИСКЛЮЧАЮЩЕЕ ИЛИ и оптическая операция И-НЕ. В таблице истинности оптического ИСКЛЮЧАЮЩЕГО ИЛИ фазой и амплитудой опорного пучка управляют таким образом, чтобы достигающий выходной плоскости волновой фронт от изображения входного сигнала и реконструкция равнялись друг другу по амплитуде и имели различие фаз в 180° . Это реализуется с помощью функции оптического ИСКЛЮЧАЮЩЕГО ИЛИ, осуществляемой между хранящимся в памяти входным сигналом и сравниваемым с изображением. В оптической операции И-НЕ фазами сигналов в опорных масках управляют в процессе записи таким образом, чтобы на стадии восстановления фаза давала требуемую оптическую операцию И-НЕ. Существует несколько способов выполнения этого процесса записи, детали которых описаны в литературе.

Особого внимания заслуживает корреляционный метод воздействия на динамическую решетку [17]. В данном подходе два складываемых числа в ССОК вводятся вдоль линейных каналов на вход совмещенного преобразующего корреляционного процессора. Данный корреляционный процессор выполняет линейную корреляцию. Можно показать, что для линейной корреляции выходной сигнал коррелятора при координатной модуляции входного сигнала представляет собой световое пятно, расположение которого определяется суммой расстояний, отсчитываемых от оси сумматора до расположения двух входных световых пятен. Для конкретного эксперимента (рис. 4.7) роль совмещенного преобразователя-коррелятора выполняет двумерный ПМС. Отраженный выходной сигнал ПМС вследствие его оптической нелинейности содержит соотношения совмещенного преобразования. Этот выходной сигнал с помощью светоделительной пластины (СП) проецируется на плоскость P_2 . Для того чтобы снизить величину выходного сигнала до величины его

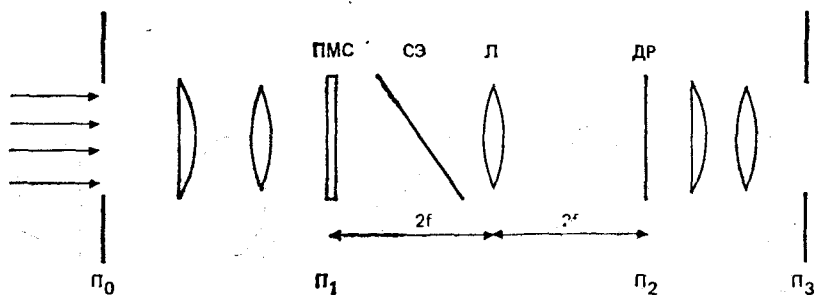


Рис. 4.7. Оптический совмещенный коррелятор-преобразователь, примененный для оптического сложения в ССОК [16].

P_0 — входная плоскость; ПМС — пространственный модулятор света; СЭ — светоделительный элемент; Л — линза; ДР — решетка (голографическая); P_1 , P_2 — промежуточная плоскость; P_3 — выходная плоскость.

периодических изменений, используется комбинация дифракционной решетки и схем управления апертурой. Пространственная частота дифракционной решетки подобрана так, что ее пространственная периодичность приспособлена для тех пространственных периодических явлений, которые хотят использовать. Эта решетка действует как пространственный фильтр в плоскости P_2 . Путем правильного управления апертурой в выходной плоскости P_3 выбирается правильный диапазон выходных чисел ССОК.

Двумерная поверхность ПМС может также быть использована для воздействия на периодическую корреляцию в отличие от предшествующей схемы управления решеткой и апертурой. Если нужно разместить входные световые пятна на окружности подходящего радиуса, тогда получающийся двумерный сигнал корреляции для фиксированных радиуса и углового направления дает круговую корреляцию. Проистекающая операция вернет ее на окружность, сообщая необходимую пространственную периодическую величину. В данном случае модульные каналы представляют собой окружности разного радиуса.

Оптические волноводные переключатели [16, 18] также являются кандидатами для оптического сложения в ССОК из-за их доступности и совершенной технологии интегральных схем (ИС). Оптические волноводные переключатели представляют собой одномодовые сильно связанные диэлектрические волноводы, чей коэффициент связи зависит через электрооптический эффект от приложенного напряжения. Вследствие того что они имеют размеры ИС, большие напряженности электрических полей (тысячи вольт на метр) могут быть достигнуты при маленьких напряжениях (порядка нескольких десятков вольт). Эти напряжения, будучи приложенными к конкретному волноводному переключателю, используются для того, чтобы перенаправить оптическое поле из одного в другой сильно связанный диэлектрический волновод. Эти волноводы работают как управляемый напряжением ответвитель (рис. 4.8, а) для оптического поля. Эти компоненты могут быть собраны в структуру, работающую как динамическая карта, где состояния электрооптических переключателей определяют конкретную, отслеживающую их карту. Существует ряд способов управлять этими переключателями, например использовать арифметические сумматоры в ССОК. Конкретный способ управления состоит в приспособлении переключающей матрицы для использования в качестве набора переключателей, управляемых напряжением. Входными сигналами матрицы являются пространственно закодированные световые пятна, это представляет первый аргумент; другим аргументом являются переключающие напряжения переключателя. Выходным результирующим сигналом переключающей матрицы является оптически пространственно закоди-

рованное световое пятно. При фиксированных напряжениях на электрооптических переключателях матрица представляет собой неподвижную карту. Неподвижные карты могут выполнять соответствующим образом прямое и обратное K -преобразование, необходимое для преобразования операции умножения в ССОК в операцию сложения в ССОК. И неподвижные, и динамические операции ССОК могут быть спроектированы и реализованы в виде модулей оптических интегральных схем (рис. 4.8, б). Фактически завершенная многозначная оптическая ИС может быть реализована в виде отдельного интегрально-оптического чипа. Основным ограничением быстродействия этого устройства является скорость включения и выключения электрооптических компонент. Эти скорости ограничены емкостными эффектами и лежат в наносекундной области.

Для того чтобы увеличить быстродействие, необходим управляемый светом оптический переключатель. В сообщении [19] описывалась управляемая светом оптическая волноводная переключающая матрица. Здесь модулирующий свет подается на волноводное устройство, имеющее фотопроводящий слой, создающий объемный заряд, который в свою очередь модулирует показатель преломления оптического волновода. В данной схеме направление распространения света обозначало различные числа в ССОК. Созданием синусоидальных дифракционных решеток на волноводных слоях можно было контролировать направление распространения волны. Возникновение поля связанного заряда является статическим эффектом, и, таким образом, можно ожидать, что такие устройства будут довольно медленными. Двумерный ПМС с перестраиваемой решеткой [30, 31] представляет собой нелинейное устройство, преобразующее локализованную интенсивность в пространственную частоту. Та-

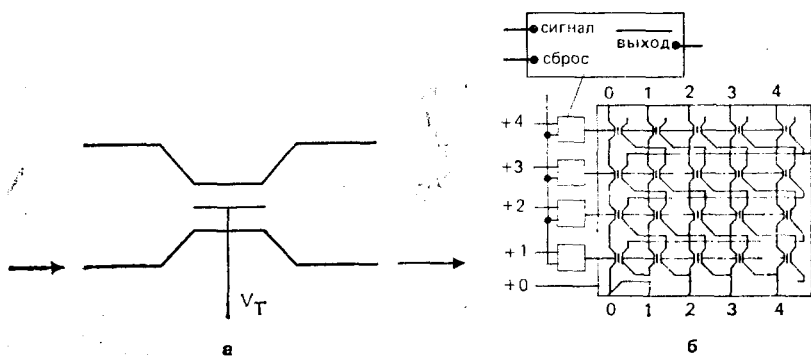


Рис. 4.8. Оптический переключатель света, управляемый напряжением: а — вариант направленного ответвителя, управляемого напряжением; V_T — управляющее напряжение; б — схема оптического сумматора, работающего по mod 5, управляемого напряжением [18].

кое устройство совместно с оптическим преобразованием Фурье может быть использовано для реализации оптических логических операций в фурье-оптике. По существу эти устройства выполняют функцию тета-модуляции, или пространственной модуляции, описанной в [24]. Фактически данная форма модуляции может быть использована для выполнения оптических операций в ССОК [25]. Недавно в ряде сообщений были описаны новые виды управляемых светом оптических переключателей. Соотношение входного-выходного сигналов переключателя зависит от конкретного нелинейного оптического эффекта. Многие из известных в настоящее время оптических переключателей представляют собой оптические резонаторы, содержащие нелинейный материал (рис. 4.9) [26, 27]. Одним из ранних оптических бистабильных устройств был резонатор Фабри—Перо, содержащий между своими зеркалами оптический нелинейный материал (рис. 4.9, а). Показатель преломления материала также зависит от интенсивности света, распространяющегося в материале. Управляя интенсивностью света в материале, можно управлять резонансной частотой резонатора. Соответственно может быть изменена интенсивность выходного сигнала

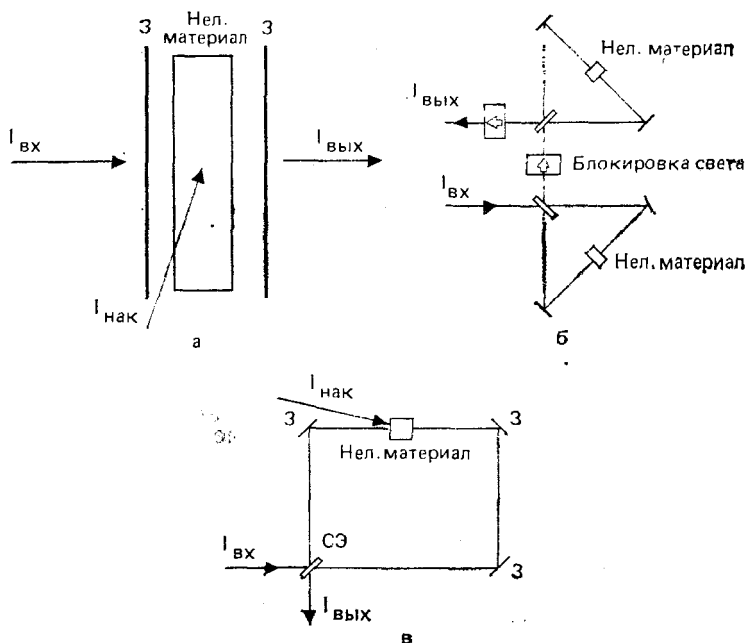


Рис. 4.9. Интерферометрические переключатели света с оптическим управлением: а — переключатель света на нелинейном резонаторе Фабри — Перо [26]; б — переключатель света на интерферометре Саньюка с накачкой сзади [27]; в — переключатель на интерферометре Саньюка с накачкой спереди [28].

ла структуры. Кроме указанного устройства, могут быть предложены другие варианты резонансных структур. На рис. 4.9, б показан оптический интерферометр на основе интерферометра Саньяка с обратной накачкой [28], в то время как на рис. 4.9, в изображен оптический переключатель на основе интерферометра Саньяка с прямой накачкой [29]. Недавно были также описаны рассеивающие беззеркальные бистабильные оптические переключатели [32]. Эти оптически управляемые переключатели света могут служить базовыми элементами для оптических вычислительных сетей. В зависимости от материала и способа возбуждения могут быть созданы различные оптические матрицы переключателей.

4.5. Проблема полноты в многозначной логике

Цель данного раздела состоит в том, чтобы изложить теорему, описывающую полную систему элементарных логических функций с k значениями входных и выходных сигналов таким образом, чтобы непосредственно сделать наглядными алгебраические свойства. Однако предварительно необходимо изложить целый ряд новых понятий, потому что наша задача заключается не только в каком-то конкретном увеличении «величины» входного сигнала, а в рассмотрении общего случая, когда число значений на входе равно некоторому произвольному числу k .

Первое из этих понятий определяется следующим образом.

Определение 4.1

h -Арным отношением R на множестве k является подмножество R множества k^h (декартово произведение k , взятое h раз, т. е. R является множеством упорядоченных наборов из h элементов множества k). Унарные отношения на k , таким образом, являются просто подмножествами; широко известными примерами бинарных отношений являются отношения порядка и отношения эквивалентности. Множество $\{00, 11, 01, 10, 10, 01, 11, 00\}$ является 4-арным отношением и бинарным отношением на $\{0,1\} \times \{0,1\}$, которое часто фигурирует при проверке полноты в случае двоичных элементарных логических функций.

Определение 4.2

Функция $f(x_0, x_1, \dots, x_{n-1})$ сохраняет h -арное отношение R на k тогда и только тогда, когда из

$$(s_{0,1}, s_{0,2}, \dots, s_{0,h}) \in R, (s_{1,1}, s_{1,2}, \dots, s_{1,h}) \in R, \dots \\ \dots, (s_{n-1,1}, s_{n-1,2}, \dots, s_{n-1,h}) \in R$$

следует, что

$$(f(s_{0,1}, s_{1,1}, \dots, s_{n-1,1}), f(s_{0,2}, s_{1,2}, \dots, s_{n-1,2}), \dots \\ \dots, f(s_{0,h}, s_{1,h}, \dots, s_{n-1,h})) \in R.$$

Если аккуратно проследить за индексами, то можно увидеть, что определение эквивалентно такому утверждению: если задана матрица с h строками и n столбцами элементов k , такими, что каждый столбец является элементом R , тогда f , определенная для каждой строки, должна дать значение столбца, являющегося элементом R . Теперь рассмотрим случай самодвойственности для двоичных функций. (Рассмотрим многозначный случай в рамках определения соответствующих отношений.) Положим f функцией трех переменных, как определено в табл. 4.1 (см. разд. 4.2). Положим, что, так же как и ранее, R будет бинарным отношением на $\{0,1\} \times \{0,1\}$ и 4-арным отношением на $\{0,1\}$. Тогда образуем следующие матрицы.

Пример 4.6

	f		f		f
101	1	100	0	001	0
110	1	010	0	001	0
010	0	011	1	110	1
001	0	101	1	110	1

Можно проверить, что столбец для f всегда дает элемент R , и, таким образом, определить, что f , в самом деле, сохраняет R . Закономерно, что f не обладает сильной полнотой. Теперь приступим к определению шести типов отношений, представляющих интерес в случае многозначной логики.

1. *Отношения порядка* с минимальным и максимальным элементом. Бинарное отношение R является отношением порядка, если оно обладает рефлексивностью, транзитивностью и антисимметрично. Рефлексивность означает $(x, x) \in R$ для всех $x \in k$. Транзитивность означает, что из $(x, y) \in R$ и $(y, z) \in R$ следует $(x, z) \in R$ для всех $x, y, z \in R$. Антисимметричность означает, что из $(x, y) \in R$ и $(y, x) \in R$ следует $x = y$.

Пример 4.7

Рассмотрим случай четырехуровневой логики на множестве $s = \{a, b, c, d\}$. Тогда следующее множество $R = \{(x_1, x_2)\}$ является отношением порядка на s :

$$R = \left\{ \begin{array}{l} x_1 = a \ b \ c \ d \ a \ a \ a \ b \ b \ c \\ x_2 = a \ b \ c \ d \ b \ c \ d \ c \ d \ d \end{array} \right\},$$

или, более кратко, $a > b > c > d$.

2. *Бинарные отношения* вида $(x, P(x))$, где P является перестановкой k с периодом k/p , где p — длина, причем p является простым числом.

Пример 4.8

Рассмотрим случай шестиуровневой логики на множестве $\{a, b, c, d, e, f\}$ и перестановки

Таблица 4.6. Коммутативная группа G , изоморфная к $Z_2 \times Z_2 \times Z_2$

+	a	b	c	d	e	f	g	h
a	a	b	c	d	e	f	g	h
b	b	a	d	c	f	e	h	g
c	c	d	a	b	g	h	e	f
d	d	c	b	a	h	g	f	e
e	e	f	g	h	a	b	c	d
f	f	e	h	g	b	a	d	c
g	g	h	e	f	c	d	a	b
h	h	g	h	e	d	c	b	a

$$\left\{ \begin{array}{l} a \ b \ c \ d \ e \ f \\ b \ c \ a \ e \ f \ d \end{array} \right\}$$

Результатом явится

$$R = \left\{ \begin{array}{l} x_1 \ a \ b \ c \ d \ e \ f \\ x_2 \ b \ c \ a \ e \ f \ d \end{array} \right\}$$

3. Если $k = p^m$ для ряда m и p , являющихся простыми числами, тогда определим 4-арные отношения для выражения

$$R = \{(x_1, x_2, x_3, x_4) \in S^4 \mid s_1 + s_2 = s_3 + s_4\},$$

где $S, +$ является коммутативной группой, в которой все элементы имеют порядок p , или p -элементарной абелевой группой. (Напомним, что группой называется множество с определенной на элементах этого множества бинарной операцией, такой, что она является замкнутой, содержит единичный элемент 1 и для каждого элемента имеется обратный член. Коммутативная группа — это группа, в которой $ab = ba$ для a, b — из группы. p -Элементарная группа — это группа, являющаяся изоморфной к $Z_p \times Z_p \times \dots \times Z_p$, где Z_p является периодической группой порядка p .)

Пример 4.9

Рассмотрим случай многозначной логики, имеющей восемь значений из множества $\{a, b, c, d, e, f, g, h\}$, и рассмотрим коммутативную группу G , определенную табл. 4.6. Эта группа изоморфна к $Z_2 \times Z_2 \times Z_2$ и фактически была образована из нее при естественной идентификации букв. Данная группа образует 4-арное отношение при установленной ранее процедуре сложения; ниже показано, как образовать это отношение.

Таблица 4.7. Многозначная таблица логических значений, обсуждаемая в примере 4.12.

$0 = 0 + 3 \cdot 0$
$1 = 1 + 3 \cdot 0$
$2 = 2 + 3 \cdot 0$
$3 = 0 + 3 \cdot 1$
$4 = 1 + 3 \cdot 1$
$5 = 2 + 3 \cdot 1$
$6 = 0 + 3 \cdot 2$
$7 = 1 + 3 \cdot 2$
$8 = 2 + 3 \cdot 2$

В отношении имеются 4 члена:

$$\left. \begin{array}{l}
 a + a = a + a \\
 b + b = b + b \\
 c + c = c + c \\
 \vdots \\
 h + h = h + h \\
 a + b = b + a \\
 a + b = c + d \\
 a + b = d + c \\
 \vdots \\
 \vdots \\
 a + b = h + g \\
 c + d = e + f \\
 \vdots \\
 \vdots \\
 a + c = c + a \\
 \vdots \\
 \vdots \\
 a + h = h + a \\
 \vdots \\
 \vdots \\
 g + b = f + c
 \end{array} \right\}$$

Данный набор содержит все возможные $x_1 + x_2 = x_3 + x_4$, заданные в табл. 4.6. Каждое уравнение порождает функцию (x_1, x_2, x_3, x_4) четырех переменных.

4. *Нетривиальные отношения эквивалентности.* (Напомним, что отношения эквивалентности есть отношения рефлексивные, симметричные и транзитивные. Симметрия означает, что для $(x, y) \in R$ выполняется $(y, x) \in R$.)

Пример 4.10

Рассмотрим случай многозначной логики, определенной на множестве шести элементов $S = \{a, b, c, d, e, f\}$. Тогда отношением эквивалентности является следующее отношение:

$$R = \{(a, c); (c, e); (a, e); (e, a); (c, a); (a, a); (c, c); (e, e); (b, d); (d, f); (b, f); (f, b); (f, d); (d, b); (b, b); (d, d); (f, f)\}.$$

5. *Центральное отношение на k .* [h -Арное отношение является центральным, тогда и только тогда, когда оно строго рефлекс-

сивно, строго симметрично и его центр не является ни пустым, ни самим множеством k . Отношение R является строго рефлексивным, если оно содержит все h членов в (s_1, s_2, \dots, s_h) , таких, что при некоторых $i \neq j$ две компоненты равны, т. е. $s_i = s_j$. (Заметим, что R может содержать много других элементов, но оно должно включать по меньшей мере эти.) h -Арное отношение является строго симметричным, если для $(s_1, s_2, \dots, s_h) \in R$ выполняется $(s_{p(1)}, s_{p(2)}, \dots, s_{p(h)}) \in R$ для каждой перестановки P , действующей на $\{1, \dots, h\}$. Центр h -арного отношения R на k определяется как $c = \{s \mid (s_1, s_2, \dots, s_{h-1}, s) \in R \text{ и } (s_1, s_2, \dots, \dots, s_{h-1}) \in k^{h-1}\}$. Заметим, что любое единичное отношение на k , не являющееся пустым и не содержащее само k , является центральным.]

Пример 4.11

Рассмотрим случай трехуровневой логики на множестве $\{a, b, c\}$. Бинарное отношение $R = \{(a, a), (b, b), (c, c), (a, b), (b, a), (a, c), (c, a)\}$ является центральным. Является ли оно строго рефлексивным? Да, $(a, a), (b, b), (c, c) \in R$. Строго симметричным? Для бинарного отношения имеется только одна перестановка, отличная от тождественной, а именно $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Непосредственно проверяется, что $(a, b), (b, a) \in R$ и $(a, c), (c, a) \in R$. В заключение вычислим центр. Справедливо ли $a \in C$? Да, поскольку $(a, a), (b, a), (c, a) \in R$. Справедливо ли $b \in C$? Нет, (c, b) не принадлежит множеству. Справедливо ли $c \in C$? Нет, (b, c) не принадлежит множеству. Следовательно, $C = \{a\}$, которое не является ни пустым, ни k , откуда R является центральным.

6. Любое отношение λ_T , определяемое h -упорядоченным классом T эквивалентных отношений на k ($s \ h > 2$). (Класс эквивалентных отношений $T = \{\theta_0, \theta_1, \dots, \theta_{m-1}\}$ является h -упорядоченным, если он удовлетворяет двум условиям. Во-первых, каждое θ_j должно иметь h классов эквивалентности ($0 \leq j \leq m-1$). Во-вторых, оно должно представлять тот случай, что пересечение $\bigcap_{j=0}^{m-1} \varepsilon_j$, где $\varepsilon_j \in \theta_j$, не является пустым (ε_j выбирают произвольно). Отношение, определяемое T , представляет собой отношение λ_T , состоящее из всех $(a_0, a_1, \dots, a_{h-1}) \in k^h$, обладающих тем свойством, что по крайней мере два элемента эквивалентны по θ_j для всех $0 \leq j \leq m-1$.)

Пример 4.12

Рассмотрим случай многозначной логики, имеющей девять уровней, т. е. $k=9$. Пусть $h=3$ и $m=2$. Согласно одной из основных теорем арифметики, любое число a_j на множестве $\{0, 1, \dots, 8\}$ может быть представлено в виде $a_j = a_{0j} + 3a_{1j}$ так, как показано в табл. 4.7. Для $i \in \{0, 1\}$ определим

$\theta_i \triangleq \{a_j \equiv a_l \text{ по } \theta_i, \text{ если } a_{ij} = a_{il}\}$

$\theta_0 = \{(0, 3, 6), (1, 4, 7), (2, 5, 8)\}$, $\theta_1 = \xi\{(0, 1, 2), (3, 5, 4), (6, 7, 8)\}$,
 $T = \{\theta_0, \theta_1\}$, λ_T является отношением на $9 \times 9 \times 9$, так что два элемента должны быть эквивалентными по θ_0 и два элемента должны быть эквивалентными по θ_1 . Список λ_T весьма длинен, и здесь будет представлено только несколько примеров. $(0, 6, 7)$, $(8, 6, 5)$, $(6, 8, 0)$ и $(6, 3, 7)$ принадлежат λ_T , в то время как $(0, 1, 2)$, $(0, 3, 6)$, $(0, 4, 8)$ и $(5, 7, 0)$ не принадлежат λ_T .

Следующая теорема принадлежит Розенбергу [22].

Теорема

Множество многозначных элементарных логических функций F является полным тогда и только тогда, когда для каждого отношения, принадлежащего одному из заданных выше шести типов, существует по крайней мере один член множества F , который не сохраняет его.

Данная теорема указывает на то, что поиск множества логически полных элементарных операций для совокупности оптических явлений должен быть в наибольшей степени связан с явлениями, разрушающими алгебраические структуры. Это находится в резком противоречии со случаем аналоговых вычислений, в которых операции сохраняют структуру в смысле линейной алгебры и линейных операторов. Детали конкретных видов алгебраических структур, которые должны быть разрушены, содержатся в шести типах отношений. Имеет ли данная теорема какой-нибудь практический смысл? И да, и нет. Она вполне полезна, если надо показать, что конкретные множества не являются полными. Для этого достаточно одного противоположного случая. Теорема также полезна для выработки подхода к вопросу, какие структуры надо разрушать.

Теорема, однако, имеет небольшое значение при определении полноты конкретного множества. Так, для пяти уровней следует проверить более 600 отношений; для 8 уровней число отношений превышает 500 миллиардов. (Подробное изложение данного вопроса можно найти в [22].) Часто для того, чтобы показать, что множество функций является полным, оказывается достаточным проявить небольшую изобретательность и сформировать из элементарных функций одну универсальную.

Логическая элементарная функция $f(x_0, \dots, x_{n-1})$ является универсальной, если она образует функционально полное множество. В данном изложении теорема Руссо интерпретируется согласно работам Дэвио. Однако следует сделать несколько предварительных замечаний. Можно рассматривать пару (k, f) , где k является произвольным множеством, в качестве алгебры. Под субалгеброй (k, f) подразумевается подмножество k' множества k , такое что $f: (k')^n \rightarrow k'$.

Дадим определение автоморфизма. Рассмотрим однозначное

отображение ϕ множества A само в себя: $A \rightarrow A$, ϕ — автоморфизм, если $f(\phi(x_0), \phi(x_1), \dots, \phi(x_{n-1})) = \phi(f(x_0, x_1, \dots, x_{n-1}))$. Наконец, нам необходимо понятие конгруэнтности на (k, f) . Пусть R_e является эквивалентным отношением на k . Тогда R_e является конгруэнтным на (k, f) , если для любого множества $\{(a_j, b_j) \mid (a_j b_j) \in R \text{ для } 0 \leq j \leq n-1\}$ справедливо $(f(a_0, a_1, \dots, a_{n-1}), f(b_0, b_1, \dots, b_{n-1})) \in R$.

Теперь можно сформулировать теорему:

Логическая функция f является универсальной на k тогда и только тогда, когда:

1. Не существует соответствующей субалгебры на (k, f) .
2. Не существует нетривиального автоморфизма на (k, f) .
3. Не существует нетривиальной конгруэнтности на (k, f) .

Две хорошо известные универсальные функции k являются функциями Уэбба и определяются так:

$$W(x, y) = \begin{cases} x + 1 \pmod k & \text{для } x = y, \\ 0, & x \neq y, \end{cases}$$

а также S -функция имеет вид

$$S(x, y) = \sup(x, y) + 1 \pmod k.$$

Доказательства не представляют большой сложности, но в случае затруднений следует обратиться к работе [2].

Другим аналогичным двоичному случаю является случай слабой полноты множества элементарных логических функций. Множество элементарных логических функций является слабо полным на k , если сложение с k постоянными функциями делает его полным. К сожалению, проверка слабой полноты требует ненамного меньшего объема работы, чем проверка полноты. Из шести заданных выше классов только второй можно полностью охватить постоянными функциями. При проверке остальных пяти классов, несомненно, не следует учитывать унарных отношений.

На этом изложение типов функций, искомых для получения полноты многозначной логики, завершено. Однако следует упомянуть раннюю теорему Слупецкого [33], так как она часто дает более простые тесты, чем теорема Розенберга. Изложение дается так, как в работе [2].

Логическая функция k переменных $f(x_0, \dots, x_{n-1})$ является существенной функцией тогда и только тогда, когда она явно зависит по крайней мере от двух переменных и принимает все значения $\{0, 1, \dots, k-1\}$. Теорема Слупецкого устанавливает, что если F является множеством логических функций и позволяет определить все унарные функции, то F является полным тогда и только тогда, когда оно содержит по крайней мере одну существенную функцию. Заметим, что это не требует по-

строения унарных функций, однако в большинстве методик это было бы принято в качестве функции памяти.

4.6. Краткое изложение и выводы

После краткого введения в вопросы полноты множеств двоичных элементарных логических функций была рассмотрена слабая полнота систем элементов, составленных из операций сложения и умножения по модулю p , являющемуся простым числом, и называемых арифметикой ССОК. Было бы разумно на базе этих компонентов непосредственно реализовать заданную переключающую функцию, хотя алгоритмы минимизации числа элементов в системе вычислений отсутствуют. Выполнение переключающих функций особенно привлекательно в ССОК благодаря широкому разнообразию методов их оптической реализации. Более того, характерной чертой почти всех оптических методов является возможность параллельной обработки в больших оптических апертурах. Этот факт указывает на огромные возможности параллельных вычислений для оптической многозначной логики. В то время как существуют аналоговые оптические методы для оптически закодированных периодических величин, таких, как фаза и поляризация, в большинстве методик оптического кодирования в качестве метода кодирования и управления модульными величинами используется пространственная координатная модуляция. Модуляция пространственного положения определяет величину динамического диапазона в области пространственных частот. Оптические системы могут достигать больших диапазонов пространственных частот. Можно рассматривать оптические многозначные логические системы как с электрической, так и с оптической адресацией. Большие достижения, полученные в последнее время в области волоконной и интегральной оптики, а также пико- и фемтосекундной оптики, показывают, что в ближайшем будущем могут стать жизненными оптические многозначные логические системы.

Последняя часть данной главы предназначена для инженеров или ученых, хорошо разбирающихся в вопросах теории и анализирующих физическую систему с позиций использования ее для построения оптических многозначных элементарных логических функций. Вопрос состоит в том, имеются ли необходимые типы входов физической системы, которые позволят синтезировать все возможные многозначные логические функции? Ответ заключается в установлении факта, сохраняет ли система определенные алгебраические структуры. Физическая система должна разрушать, не сохранять шесть классов алгебраических структур. В поисках полноты многозначных множеств в физических системах следует отыскивать системы, уничтожающие упорядоченность.

Глава 5

Пороговое кодирование и взвешивание в оптических вычислениях

*С. С. Густафсон, Исследовательский институт,
Отделение прикладной физики, Университет
г. Дайтона, Дайтон, шт. Огайо*

5.1. Введение

В данной главе дается обзор возможностей оптической техники для выполнения процедур принятия элементарных решений и операций межэлементных соединений, необходимых в любых системах обработки информации. В разд. 5.1 рассматриваются случаи, в которых оптически выполняются только операции соединений; в разд. 5.3 приводятся примеры устройств, в которых и соединения, и операции принятия решения выполняются оптическими методами; в разд. 5.4 обсуждение основных проблем завершается описанием общей схемы оптического вычислительного устройства.

Поскольку оптические вычисления являются сравнительно новой областью, то представляется целесообразным использовать систему счисления в остаточных классах (ССОК), многозначную и пороговую логику и связанные с ними архитектуру и программные средства, не нашедшие широкого применения в чисто электронных вычислительных машинах. В целом исследование указанных нестандартных средств достигло в настоящий момент лишь такого уровня, при котором могут быть сделаны не более чем предварительные оценки их значимости для оптических вычислений. Данное утверждение одинаково справедливо и для операций порогового кодирования в оптических вычислениях, и для процедур взвешивания, описанных в разд. 5.1.2. В данной главе вопросы излагаются по возможности кратко и абстрактно, а все примеры просты и в основном мотивируют будущие исследования.

5.1.1. Возможность оптических вычислений

Оптические вычисления обладают значительными возможностями улучшения (на несколько порядков) таких характеристик, как быстродействие, потребление энергии, размер, объем памяти, надежность, отказоустойчивость и т. д. по сравнению с существующими чисто электронными вычислительными системами. (Заметим, что улучшение нескольких характеристик од-

новременно может быть практически нецелесообразным.) Принципиальным преимуществом оптических методов (характеризуемых диапазоном возможных рабочих частот до 10^{14} Гц) по сравнению с электронными методами (характеризуемыми возможными рабочими частотами до 10^7 Гц) является их способность выполнять сравнительно многочисленные, сложные, глобальные и широкополосные соединения [1, 2]. Такие оптические соединения в целом отличаются (а) отсутствием интерференционных эффектов (т. е. оптические пучки могут пересекаться, не вызывая явления интерференции, а электронные «провода» — не могут) и (б) относительной независимостью от свойств линии связи, которые могут привести к значительным временным задержкам и соответствующим ограничениям рабочих характеристик. Другим возможным преимуществом оптических методов является параллельная обработка массивов данных, высокое быстродействие логических элементов и простая реализация трехмерных конструкций. Большинство из указанных свойств могут быть присущи и чисто электронным методам и, таким образом, носят менее принципиальный характер по сравнению с преимуществами при реализации соединений. В целом оптические методы имеют превосходство при выполнении операций соединения и по крайней мере равные возможности с чисто электронными технологиями при выполнении операций принятия решений. Это различие может быть связано с тем фактом, что степень нелинейности материала, необходимая для операций принятия решения, гораздо сильнее проявляется при частотах и интенсивностях, присущих чисто электронным методикам.

Возможности оптических вычислений можно оценить, например, при рассмотрении свойств системы, состоящей из матрицы оптических бистабильных устройств с чисто оптической цепью обратной связи. В данном случае бистабильное устройство выполняет операции принятия решения, а цепь обратной связи (которая может содержать голограммы, линзы, светоделительные элементы и другие компоненты) выполняет операции соединения. Как показано в разд. 5.4, обоснованные планы развития современной технологии могут обеспечить выполнение такими системами порядка 10^{15} логических операций в секунду на площади в 1 см^2 , или более чем в 100 раз выше, чем планируют современные программы по СБИС [3]. Существующее ограничение характеристик этой системы состоит в рассеянии мощности матрицей оптических бистабильных элементов; данное ограничение является свидетельством относительной сложности использования оптики для процедур принятия решения. Однако эта трудность не должна сдерживать развитие оптических компьютеров с феноменальными рабочими характеристиками по сравнению с чисто электронными вы-

числительными машинами. Например, многочисленные и сложные соединения, в которых оптика имеет превосходство, возможно, являются наиболее важной чертой архитектур биологических нейронных сетей, в которых каждый логический элемент (или нейрон) может быть соединен с приблизительно 10^4 других элементов [4, 5]. Хотя время переключения или время принятия решения отдельного нейрона велико по сравнению с характерными временами переключения электронных вентилях, широко известны исключительные возможности этих биологических архитектур (например, при распознавании образов с высокой степенью адаптивности в миниатюрных и потребляющих мало энергии «живых» структурах).

5.1.2. Определения взвешивания и порогового кодирования

Архитектура компьютера и выполняемые им операции могут быть представлены в виде набора соединенных между собой элементов, которые принимают решения. Например, любая булева логическая функция (или таблица истинности) может быть реализована цепью элементарных решателей или логических элементов, таких, как вентили И или ИЛИ. В целом соединения могут быть охарактеризованы понятием веса, численно описывающего состояния межэлементного соединения. Например, состояния отсоединения и подсоединения могут быть описаны нулем (0) и единицей (1), либо даже любым действительным или комплексным числом (возможные значения комплексных весовых коэффициентов рассматриваются в разд. 5.2.1). Решатели могут характеризоваться неравенствами, связанными с множеством пороговых значений, которые также могут быть действительными или комплексными числами. Таким образом, соединения могут ассоциироваться с взвешиванием, а принятие решений — с пороговым кодированием. Следует обратить внимание на то, что взвешивание представляет собой более общее понятие, чем соединение, а принятие решения — более общий термин, чем пороговое кодирование.

Пространственные и временные характеристики системы, связанные с операциями взвешивания и порогового кодирования, могут быть использованы для классификации архитектур оптических вычислений [6]. А поскольку основное различие архитектур обычно связывают с наиболее трудными для реализации операциями, поэтому классификацию архитектур целесообразно связать именно со взвешиванием и пороговым кодированием. Соответственно к первому классу относят архитектуры с внешним пороговым кодированием, где операции порогового кодирования выполняются в неоптической части одного или более оптико-электронных интерфейсов, а операции взвешивания выполняются такими оптическими элементами, как го-

лограммы, линзы и т. д. Например, в ставшем теперь классическим стэнфордском умножителе матрицы на вектор [7] (использованном для выполнения дискретных линейных преобразований, например дискретных фурье-преобразований) в целом для реализации соединений используется маска с несколькими апертурами и пассивная оптика, а для принятия решений — используется матрица фотодетекторов с электронным пороговым кодированием. Ко второму классу относят архитектуры с внутренним пороговым кодированием, где по крайней мере часть операций порогового кодирования выполняется чисто оптически и где, следовательно, необходимы нелинейные устройства с оптическими входными и выходными сигналами. Таким образом, архитектура с внутренним пороговым кодированием содержит чисто оптические нелинейные устройства, в то время как архитектура с внешним кодированием не содержит таких элементов, и классификация отделяет операции принятия решений, выполняемые оптически, от операций, осуществляемых электронными устройствами. Указанная классификация может быть использована применительно к цифровым и аналоговым схемам или к несинхронизированным и синхронизированным архитектурам (например, комбинационным, последовательно асинхронным или последовательно синхронным). Классификация применена для различных уровней сложности устройств — от отдельных вентилях до готовых процессоров [8]. Следует обратить внимание, что архитектуры обоих классов, как правило, базируются на возможностях оптики осуществлять (с помощью, например, линз, голограмм и т. д.) соединения без интерференционных явлений.

5.1.3. Стандартная и нестандартная пороговые логики

Пороговая логика — это область, тесно связанная с операциями порогового кодирования и взвешивания в вычислениях. Эта область активно исследовалась в 1960-е и начале 1970-х гг. [9—11], но с тех пор и до настоящего времени [12] интерес к ней спал в основном потому, что общепринятой во всех чисто электронных ИС стала традиционная булева логика.

Пример стандартной, или построенной на линейных неравенствах пороговой логики наглядно показан на рис. 5.1. Здесь логическая функция $y = \overline{x_1 x_2 x_3 x_4} + x_1 x_3$ (для которой дана таблица истинности) реализуется с помощью пороговых логических элементов, у которых x_1, x_2, x_3 и x_4 — входные двоичные сигналы, y — двоичный выходной сигнал, знак «+» обозначает операцию ИЛИ, подразумеваемое умножение является операцией И, а черта — операцией НЕ. Элемент умножает каждый двоичный входной сигнал на действительное число, являющееся весовым коэффициентом (w_1, w_2, w_3 или w_4),

суммирует результаты и сравнивает их с пороговым действительным значением T . Если сумма меньше порогового значения, то выходной сигнал равняется 0, в противном случае — 1. В приведенном примере следует обратить внимание на то, что единственный пороговый элемент реализует логическую функцию, которая потребовала бы нескольких обычных булевых вентилях и двух уровней логики (трех уровней в том случае, если присутствует операция НЕ). Следует также заметить, что, во-первых, пороговый элемент будет правильно срабатывать в случае, если порог является любым числом, удовлетворяющим условию $4 < T \leq 5$, и, во-вторых, если допустимое отклонение значения порога ограничено, тогда могут быть определены ненулевые допустимые отклонения весовых коэффициентов.

В общем пороговые логические элементы с линейными неравенствами имеют несколько двоичных входов, один двоичный выход и встроенное аналоговое устройство, пригодное для работы с оптикой (некоторые конструкции элементов допускают применение дискретных встроенных устройств [13] и несколько двоичных выходов [14]). Как показано на рис. 5.1, можно добиться, чтобы пороговый элемент работал корректно даже в случае отклонения значения порога и веса от номинальных величин, что возникает вследствие внешних условий, погрешностей при изготовлении прибора и т. д. Из таблицы истинности на рис. 5.1 видно, что соответствующие величины веса и порога могут быть получены путем решения системы линейных неравенств. Например, 12-я строка таблицы истинности требует, чтобы

$$w_1x_1 + w_3x_3 + w_4x_4 \geq T. \quad (5.1)$$

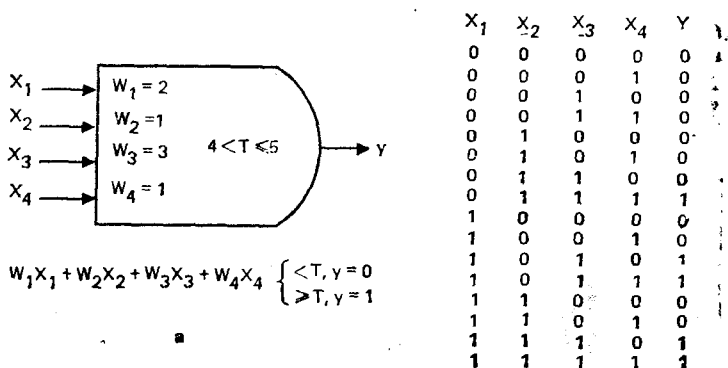


Рис. 5.1. Пример порогового логического элемента, выполняющего функцию $y = \bar{x}_1x_2x_3x_4 + x_1x_3$: а — пороговый элемент; б — таблица истинности.

Обычно число неравенств превышает число неизвестных (в приведенном примере 16 неравенств, или строк таблицы истинности приходится на 5 неизвестных $\omega_1, \omega_2, \omega_3, \omega_4$ и T), однако решения будут существовать не всегда. Тем не менее существующие решения, определяющие максимальные допустимые отклонения различных порогов и весов, могут быть определены с помощью методов линейного программирования. Большинство ранних работ по пороговой логике было посвящено нахождению таких решений или их характеристик без явного применения методов линейного программирования, часто требующих проведения обширных вычислений. Совершенствование вычислительной техники и алгоритмов, однако, сделало использование линейного программирования для разработки пороговых логических элементов и схем более привлекательным [15].

Как показано выше, не все логические функции могут быть реализованы с помощью пороговых логических элементов с одним линейным неравенством; те из них, которые могут быть реализованы, называются пороговыми, или линейно-разделяемыми функциями. В целом существует 2^{2^n} логических функций n двоичных переменных (каждая из 2^n входных строк таблицы истинности может иметь любой двоичный выход), но число пороговых функций обычно намного меньше — верхний предел их числа составляет $(2^n + 1)/n!$. Например, если $n=3$, полное число функций равняется 256, верхнее предельное значение составляет 170, а фактическое число функций оказывается равным 104 [10]. Для $n=2$ (простейший случай) можно легко показать, что 14 из 16 возможных булевых логических вентилей с двумя входами, включая И и ИЛИ, могут быть реализованы с помощью единственного порогового элемента; таким образом, линейные неравенства пороговой логики можно рассматривать как более общий случай булевой логики. Поскольку любые комбинационные логические функции (с таблицей истинности из постоянных значений) можно реализовать на основе системы вентилей или элементов с не более чем двумя уровнями булевой логики (т. е. сигнал в системе не должен проходить более двух последовательно соединенных логических вентилей, исключая вентиль НЕ), то оказывается, что то же самое справедливо для пороговой логики. Однако булевы логические схемы для сложных функций (например, 16-разрядный умножитель) обычно требуют более двух логических уровней, чтобы избежать соединений на одном и том же уровне неоправданно большого числа логических элементов [16]. Пороговая логика, в частности реализация пороговой логики в оптике, может смягчить эти требования. Данная характеристика и пример на рис. 5.1 показывают, что пороговая логика имеет потенциальные преимущества, обеспечивая мень-

шее число логических уровней (что приводит к большему быстродействию), логических элементов (что уменьшает энергопотребление) и соединений (что уменьшает сложность системы).

Данные преимущества обычно сильнее выражены для более общих случаев нестандартных типов пороговой логики, или логик с нелинейными неравенствами. Элемент в подобной логической схеме производит сравнение пороговой величины с нелинейной функцией двоичного входного сигнала, такой как квадратичный полином, коэффициенты которого рассматривают в качестве весовых коэффициентов [17]. Как утверждается в разд. 5.2, когерентные оптические системы реализуют нелинейную функцию пропорционально возведенной в квадрат сумме двоичных входных сигналов, умноженных на комплексные значения весовых множителей, где последние представляют амплитуды и фазы световой волны. С помощью исчерпывающего, но имеющего ограниченную точность численного расчета было показано, что число логических функций при $n=3$ двоичных переменных для случая одного порогового элемента с комплексными весами и как инвертированными, так и неинвертированными выходами составляет по меньшей мере 246 [18]. По сравнению со 104 такими функциями, которые реализуются с помощью стандартной пороговой логики, это число показывает увеличение вычислительных возможностей одного из вариантов нестандартной пороговой логики. Некоторые архитектуры оптических вычислительных схем [4, 5, 19, 20] описываются с помощью общепринятых систем пороговой логики, а в случае использования на практике их рабочие характеристики могут быть улучшены. Например, если число разновидностей n двоичных элементов, которые необходимо классифицировать по одному или двум признакам, составляет величину менее $2n$, тогда хорошо известные адаптивные методы, основанные на алгоритме Видрова — Хоффа или на архитектуре персептрона, обычно могут идентифицировать линейно-разделяемую функцию или отдельный стандартный пороговый элемент, выполняющий классификацию [10]. Однако при наличии возможности использования одиночного нелинейного порогового элемента (имеющего эффективное приложение в оптике) можно надлежащим образом классифицировать более чем $2n$ разновидностей элементов [21].

5.2. Устройства с внешним пороговым кодированием

В устройствах с внешним пороговым кодированием оптические средства используются только при реализации соединений или операций взвешивания; принятие решения или операция порогового кодирования выполняются электронными или опто-

электронными методами (например, фотодетекторами с пороговыми усилителями). Операции взвешивания обычно осуществляются с помощью набора дифрагирующих, преломляющих и отражающих свет элементов в обычных или интегрально-оптических средствах. Широко распространенные дифракционные решетки или голограммы являются основными компонентами в вариантах устройств, в которых для получения желаемых значений весовых коэффициентов используются концепции запоминающих устройств с выборкой по содержанию или методик голографических таблиц перекодировки [22]. В разделах 5.2.1 и 5.2.2 рассматриваются два простых примера этих моделей: конструкция 2-разрядного умножителя и конструкция, реализующая любую логическую функцию двух переменных.

5.2.1. Пример двухразрядного умножителя

Таблица истинности для умножителя двух 2-разрядных чисел x_1x_0 и y_1y_0 с результатом $z_3z_2z_1z_0$ представлена на рис. 5.2. Предположим, что четыре входных бита представлены 0, если они сами нулевые, и величинами $x_1 = A_1 \exp(i\phi_1)$, $x_0 = A_2 \exp(i\phi_2)$, $y_1 = A_3 \exp(i\phi_3)$ и $y_0 = A_4 \exp(i\phi_4)$, если входные сигналы равны 1. Далее, если эти выражения описывают волны, и мы предполагаем, что все расстояния источник — источник, источник — детектор и детектор — детектор велики по сравнению с длиной волны, то конструкция двухразрядного умножителя может выглядеть так, как показано на рис. 5.3, а [6]. Здесь x_0 , x_1 , y_0 и y_1 являются точечными источниками света, а z_0 , z_1 , z_2 и z_3 являются точечными фотодетекторами. Линии указывают оптические пути, каждый из которых может иметь выбранный коэффициент затухания и фазовый сдвиг, что реализуется с помощью голограмм или дифракционных элементов интегральной оптики.

Требуемые коэффициенты затухания и фазовые сдвиги определяются путем решения систем нелинейных неравенств, получаемых из таблицы истинности. Например, 12-я строка и столбец z_2 таблицы (обведено в рамку на рис. 5.2, а) дают сигнал фотодетектора z_2 , который должен быть больше или равен порогу T_2 :

$$|A_1 \exp(i\phi_1) + A_3 \exp(i\phi_3) + A_4 \exp(i\phi_4)|^2 \geq T_2. \quad (5.2)$$

Аналогичные выражения могут быть получены таким образом, что каждый из четырех столбцов выходного сигнала (обозначенных z_3 , z_2 , z_1 и z_0) описывается системой из 16 нелинейных неравенств (по одному на каждую строку таблицы). Неравенства являются функциями 9 переменных: четырех амплитуд (A_1 , A_2 , A_3 и A_4), четырех фаз (ϕ_1 , ϕ_2 , ϕ_3 и ϕ_4) и одного порога (T_1 , T_2 , T_3 или T_4). Требуется найти решение для каждой

X_1	X_0	Y_1	Y_0	Z_3	Z_2	Z_1	Z_0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	0
0	0	1	1	0	0	0	0
0	1	0	0	0	0	0	0
0	1	0	1	0	0	0	1
0	1	1	0	0	0	1	0
0	1	1	1	0	0	1	1
1	0	0	0	0	0	0	0
1	0	0	1	0	0	1	0
1	0	1	0	0	1	0	0
1	0	1	1	0	1	1	0
1	1	0	0	0	0	0	0
1	1	0	1	0	0	1	1
1	1	1	0	0	1	1	0
1	1	1	1	1	0	0	1

$$б \quad \left| A_1 e^{i\theta_1} + A_3 e^{i\theta_3} + A_4 e^{i\theta_4} \right|^2 \geq T$$

Рис. 5.2. а — таблица истинности двухразрядного умножителя; б — нелинейное неравенство для помещенных в рамку входных значений в таблице истинности.

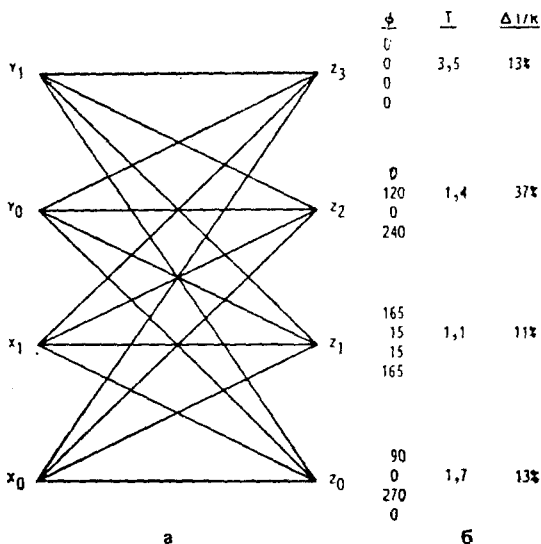
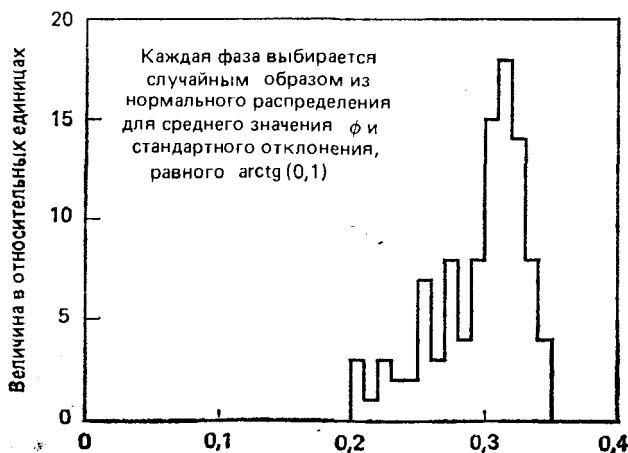


Рис. 5.3. Двухразрядный умножитель с внешним пороговым кодированием: а — соединения источников x и y с детекторами z ; б — решения (без учета затухания) для фаз соединений ϕ , порогов детектирования T и отклонений порога $\Delta T/R$.

из четырех преобразованных систем неравенств, в которую входят такие члены, как A_1^2 , $2A_1A_2 \cos(\phi_1 - \phi_2)$ и т. д., и которая содержит только восемь неизвестных (так как, например, независимую переменную $\phi_1 - \phi_2$ можно заменить либо на ϕ_1 , либо на ϕ_2). Все полученные величины амплитуд, фаз и порогов должны иметь допустимые отклонения или диапазоны, в которых они могут изменяться, не влияя на корректную работу 2-разрядного умножителя.

На рис. 5.3, б представлено одно из решений, которое включает только фазовый сдвиг (затухание волны отсутствует) и может иметь приемлемое на практике допустимое отклонение [6]. Справа на рисунке в столбце ϕ представлены значения фазовых сдвигов, необходимых для каждого из четырех расстояний до каждого из детекторов, столбец T дает значения порога для каждого детектора при $A_1 = A_2 = A_3 = A_4 = 1$, а столбец $\Delta T/R$ показывает долю 'полного диапазона изменения сигнала для каждого из детекторов, за пределами которого порог может изменяться. На рис. 5.4 показана гистограмма $\Delta T/R$ для выходного разряда z_2 , получаемого путем случайного выбора каждой из четырех фаз для этого выходного сигнала из нормальных распределений со средними значениями выбранных величин и стандартным отклонением, равным $\arctg(0,1)$. Эти стандартные отклонения соответствуют 10% смещению фазовых векторов, и на рис. 5.4 показано, что такие отклонения снижают допустимое отклонение порога $\Delta T/R$ для выходного разряда z_2 от 37% до приблизительно 20%. Аналогичные величины допустимых отклонений могут быть получены для других выходных разрядов.



Гистограмма отклонения порога $\Delta T/R$ для выходного значения z_2

Рис. 5.4. Гистограмма отклонений порога для выходного сигнала z_2 .

Описанная выше схема 2-разрядного умножителя основана на способности света осуществлять безынтерференционные соединения, которые (а) являются параллельными в том смысле, что время соединения существенно не зависит от длины соединения или его веса и (б) приводят к временам срабатывания системы, существенно ограниченным лишь только временем отклика источников и детекторов. Эти соединения могут быть реализованы с помощью пассивных дифрагирующих элементов в виде одной или более голограмм. Могут использоваться обычные «тонкие» голограммы, «объемные» голограммы, трехмерные тонкослойные, а также трехмерные голограммы с глубокой записью, в которых свет когерентный или, возможно, некогерентный (белый) [23] распространяется приблизительно по нормали к плоскости голограммы. (В устройствах с комплексными весовыми коэффициентами возможно использование некогерентного света, например, если с помощью «объемной» голограммы не только осуществляют соединения, но также выделяют требуемые для восстановления изображения длины волн и фазы некогерентного света, что является типичным для голограмм, восстанавливаемых в дневном свете.) Эти соединения могут также быть осуществлены в интегрально-оптических устройствах с помощью пассивных дифрагирующих элементов, образованных на поверхности подложки либо вблизи нее, так что свет распространяется приблизительно параллельно поверхности. Такие интегрально-оптические устройства могли бы использовать поверхностный рельеф или фоторефрактивные механизмы для получения дифракционных элементов на GaAs, LiNbO₃, стекле или других подложках. Данные устройства обладают потенциальными преимуществами с точки зрения размеров, энергопотребления, надежности, а также программируемых в реальном времени соединений весовых коэффициентов. Для реализации этих преимуществ следует использовать структуры дифрагирующих элементов с электронной модуляцией [6, 24].

Созданные оптическими методами голограммы могут реализовывать процедуры взвешивания, необходимые согласно заданным таблицам истинности или отношениям входного-выходного сигналов (включая случай 2-разрядного умножителя) в системах с внешним пороговым кодированием (хотя системы с большим числом входов могут оказаться невыгодными на практике). Используя стандартные модели голографических процессов [25], можно показать [6], что выходная матрица истинности A , имеющая $L \times P$ элементов, связана с входной матрицей истинности C , имеющей размерность $N \times P$, соотношением

$$A = OR^+C. \quad (5.3)$$

Здесь O и R — матрицы размерностей $L \times M$ и $N \times M$, соответ-

ственно описывающие комплексные амплитуды, использованные для записи M раз экспонированных голограмм, и \dagger обозначает операцию сопряженного транспонирования матрицы. Важной особенностью уравнения (3) является то, что, хотя для записи голограмм может быть использована многократная экспозиция, способность голограмм содержать в себе отношения входного-выходного сигналов описывается не более чем NL комплексными элементами матрицы OR^\dagger . В 2-разрядном умножителе, например, где $N=L=4$ и $P=16$, только 16 комплексных параметров могут быть использованы для связи 64 битов входного сигнала с 64 битами выходного сигнала. Это предполагает, что в записанной оптически голограмме не могут быть реализованы все возможные таблицы истинности. Аналогичная ситуация (обсуждавшаяся в разд. 5.1.3) состоит в том, что не все логические функции могут быть реализованы одиночными пороговыми логическими элементами.

Было бы полезно решить уравнение (5.3) по крайней мере для OR^\dagger , выражая ее элементы через C и A . Это матричное уравнение обычно преобразовывают, а для получения приближенного решения используется метод наименьших квадратов или псевдоинверсии. Решение по методу наименьших квадратов (строка за строкой), например, дает

$$OR^\dagger = AC^\dagger(CC^\dagger)^{-1}. \quad (5.4)$$

И хотя данное решение может не давать необходимой таблицы истинности для системы с внешним пороговым кодированием, однако оно в любом случае служит отправной точкой, например, для применения метода «скорейшего спуска» или других вычислительных методов поиска необходимых решений. Такие решения в случае, если элементы матрицы изменяются в определенных диапазонах отклонений, должны поддерживать требуемое отношение входного-выходного сигналов. Изобразитель с внешним пороговым кодированием является таким решением и используется для определения матрицы OR^\dagger , в которой все элементы имеют единичные величины [6]:

$$OR^\dagger = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \frac{(-1 + \sqrt{3}i)}{2} & 1 & \frac{(-1 - \sqrt{3}i)}{2} \\ \frac{(-\sqrt{15} + i)}{4} & \frac{(\sqrt{15} + i)}{4} & \frac{(\sqrt{15} + i)}{4} & \frac{(-\sqrt{15} + i)}{4} \\ i & 1 & -i & 1 \end{bmatrix}. \quad (5.5)$$

Конкретной реализацией данного решения для случая голографической записи являются тождественная матрица O размерности 4×4 и матрица $R = (OR^\dagger)^\dagger$. Следует обратить вни-

мание на то, что, хотя вышеприведенный анализ подразумевает трехмерные голографические системы, могут быть использованы интегрально-оптические сборки дифракционных элементов, близкие по своим функциям к объемным голограммам. Данная возможность связана с тем фактом, что ряд «образов» таблицы истинности, которые надо записать и восстановить, могут представлять собой сравнительно простые изображения, не требующие высокого разрешения и состоящие из светлых и темных пятен, хотя при этом зачастую и обладают сильной перекрестной корреляцией.

5.2.2. Пример 2-разрядного перестраиваемого устройства

Взвешивание и соединение, осуществляемые оптически с помощью компьютерно синтезированных голограмм, обычно требуют знания дифракционных картин голограммы. При этом голограммы, изменяющие амплитуды и фазы, позволяют получить корректную таблицу истинности, описывающую соотношение входного-выходного сигналов с максимальным весовым коэффициентом и отклонением порога. В случае геометрической оптической схемы 2-разрядного умножителя, показанного на рис. 5.3, были легко получены выражения, определяющие поведение соотношения входного-выходного сигналов. Данная благоприятная ситуация может являться исключением при конструировании систем с внешним пороговым кодированием, как правило, имеющих меньшие размеры, большую эффективность и т. д., для которых приближение геометрической оптики не применимо.

Рассмотрим, например, получение восьми голограмм дальнего поля (каждая из них для тех же параметров, что у двух указанных выше разработок), которые осуществляют в системе с внешним пороговым кодированием восемь функций двух булевых переменных, имеющих положительный порог (например, восемь из тех шестнадцати функций, для которых два нулевых входных сигнала порождают нулевой выходной сигнал). На рис. 5.5 показана простая оптическая схема, состоящая из экрана с двумя маленькими отверстиями, разнесенными на расстояние y . Одно из отверстий покрыто пленкой θ , изменяющей фазу; имеется детектор d и верхний и нижний взаимно когерентные точечные источники l и u . В приближении дальнего поля расстояния b и y , а также длина волны $\lambda = 2\pi/k$ должны быть малы по сравнению с расстоянием z . В рамках данного приближения и при фиксированном b задача сводится к нахождению величин y и θ , таких, что протектированный сигнал I_l только для включенного источника l , сигнал I_u , полученный только при включенном источнике u , и сигнал I_b , полученный при обоих включенных источниках, имеют все шесть

возможных неравенств. Обращаясь за определениями к рис. 5.5, получаем следующие приближенные соотношения:

$$\begin{aligned}
 A_l &= \exp i [k(\omega + s)] + \exp i [k(v + r) + \theta], \\
 A_u &= \exp i [k(\omega + s)] + \exp i [k(u + r) + \theta], \\
 I_l &= |A_l|^2 \approx (ks)^2 (x^2 + ax)^2 + 2\eta(ks)(x^2 + ax) + \eta^2, \\
 I_u &= |A_u|^2 \approx (ks)^2 (x^2 - ax)^2 + 2\eta(ks)(x^2 + ax) + \eta^2, \\
 I_b &= |A_l + A_u|^2 \approx 2(ks)^2 [(x^2 + ax)^2 + (x^2 - ax)^2] + \\
 &\quad + 8\eta(ks)x^2 + 4\eta^2 - 4(ks)^2(ax)^2,
 \end{aligned} \tag{5.6}$$

где $x = y/s$, $a = b/s$ и $\eta = \varphi - \pi \approx 0$. На рис. 5.6 показаны графики [6] приближенных выражений для I_l , I_u , I_b в зависимости

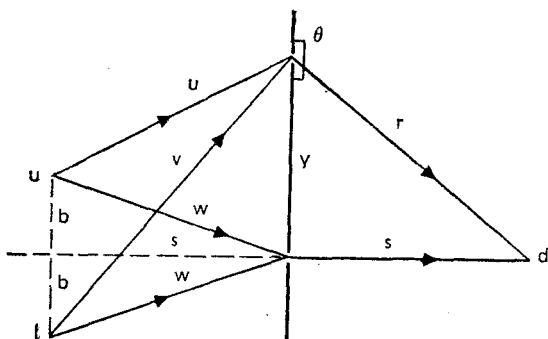


Рис. 5.5. Схема получения голограмм дальнего поля, реализующих все восемь положительных пороговых функций двух булевых переменных в системе с внешним пороговым кодированием.

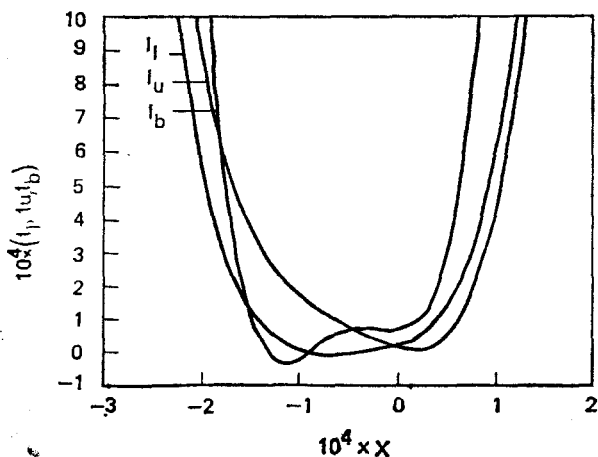


Рис. 5.6. Графики приближенных зависимостей I_l , I_u и I_b от x при $\lambda = 628$ нм, $b = 10$ мкм, $s = 10$ см и $\eta = 0,004$.

от x для $\lambda=628$ нм, $b=10$ мкм, $s=10$ см и $\eta=0,004$. Заметим, что четыре из шести неравенств могут быть удовлетворены при указанных значениях; два других соотношения могут быть удовлетворены для других величин η .

Пример уравнения (5.6) в целом демонстрирует возможную сложность задачи внешнего порогового кодирования (в рамках физической оптики). Большую часть трудностей можно было бы предсказать, если удовлетворить условиям дифракции Фраунгофера и если таблицы истинности, описывающие соотношение входного-выходного сигналов имеют большое число возможных корректных вариантов. Один из подходов к проблемам такого синтеза состоит в выполнении дополнительной обработки после фотодетектирования и использования методик логического упрощения и арифметики ССОК с целью уменьшения объема информации, которая должна быть накоплена для голографической перекодировки [22]. Более общий подход заключается в поиске возможных условий перехода в ССОК и обратно и способов дополнительной чисто электронной обработки.

Данный подход позволял использовать (а) получение, как правило, больших систем нелинейных кодирующих неравенств, полностью описывающих требуемую систему внешнего порогового кодирования; (б) нахождение оптимальных решений для этих систем, по возможности относительно отклонений весового коэффициента и порога, используя методики нелинейного программирования [26]; (с) идентификацию оптических систем (возможно основанных на компьютерно-синтезированных голограммах), реализующих решения. В случае оптически записанных голограмм недавняя публикация [27] показывает, что управление относительными фазами таблицы истинности кодирующих предметных пучков является весьма важным для получения избирательного восстановления без интерференции (от неселективного восстановления). В другой недавней работе было показано, что такая интерференция может быть уменьшена при использовании методик, где применяется сравнение усиления сигнала в нелинейных оптических резонаторах, содержащих, например, искажающие фазу зеркала [28]. Указанные методики явно используют оптическую процедуру принятия решения и, таким образом, относятся к категориям устройств с внутренним пороговым кодированием.

5.3. Устройства с внутренним пороговым кодированием

Устройства с внутренним пороговым кодированием используют оптические методики для операции принятия решения, так же как и для операций соединения. Следовательно, они представляют собой более общий класс устройств, чем системы с

внешним пороговым кодированием, в которых не допускаются оптические операции принятия решений. В частности, использование двух или более уровней принятия решения в принципе позволяет выполнить любую комбинаторную логическую функцию, но устройства с внешним пороговым кодированием используют лишь один уровень принятия решения и, таким образом, в принципе не обладают этой способностью. Как было указано в разд. 5.1, осуществляющие операции принятия решения оптические системы не нуждаются в применении тех схем вентиляльных устройств с булевой логикой, которые стали традиционными в чисто электронных интегральных схемах. В частности, пороговая логика утверждает более общий, включающий традиционную булеву логику в качестве частного случая подход, позволяющий уменьшить число используемых логических уровней, элементов и соединений.

Два простых примера устройств с внутренним пороговым кодированием, одно из которых (умножитель-сумматор) использует комбинаторную логику, а другое (J - K -триггер) — последовательную логику, рассмотрены в разд. 5.3.1 и 5.3.2. Эти устройства могут быть реализованы с высокой степенью интеграции на основе нелинейных оптических (бистабильных) устройств. Так как оптико-электронные (или электронно-оптические преобразования), как правило, приводят к ухудшению таких показателей, как быстродействие, энергопотребление, геометрические размеры и т. д., можно ожидать, что эти приборы потребуют чисто оптических или близких к ним внутренних соединений с целью улучшения рабочих характеристик по сравнению с чисто электронными устройствами.

5.3.1. Пример умножителя-сумматора

Рассмотрение данного примера было вызвано необходимостью обеспечить высокую скорость при выполнении операции внутреннего произведения в линейной алгебре (например, для умножения матрицы на вектор или матрицы на матрицу), в противном случае эти операции становятся бессмысленными. Операции внутреннего произведения включают умножение двух чисел и сложение результата с третьим числом. Например, 2-разрядный умножитель-сумматор умножает два 2-разрядных числа M и N , прибавляет результат к 5-разрядному входному числу X и выводит результаты в виде 5-разрядного числа Y . В синхронизированном режиме работы выходной сигнал Y мог бы подаваться по цепи обратной связи на вход X для того, чтобы достичь эффекта многократного накопления результата (если имеется возможность накопления до трех произведений и при этом не возникает переполнение).

Традиционная схема булевой логики для 2-разрядного умножителя-сумматора изображена на рис. 5.7, где индексы у M , N , X и Y обозначают двоичные номера позиций (2^0 , 2^1 и т. д.). Следует обратить внимание на то, что хорошо известные и часто встречающиеся многовходовые системы были сгруппированы и обозначены одним символом. Функция ИСКЛЮЧАЮЩЕЕ ИЛИ может быть реализована с помощью двух логических уровней с использованием двух вентилях И, одного ИЛИ и двух инверторов. Функции чисто суммирующего устройства можно выполнить при двух логических уровнях при помощи как минимум пяти вентилях И, двух вентилях ИЛИ и четырех инверторов. На рис. 5.8 изображена пороговая логическая схема 2-разрядного умножителя-сумматора [29, 30], составленная полностью из пороговых логических элементов с коэффициентами объединения по входу и разветвления по выходу, ограниченными пятью. Весовые коэффициенты указаны внутри каждого элемента символом, смежным с линией, обозначающей вход, а порог указан по соседству с выходной линией.

Следует заметить, что булева логическая схема на рис. 5.7 требует всего 38 логических вентилях и 18 инверторов. (Инверторы обычно не включают в число логических элементов или общее число логических уровней, однако они занимают объем, требуют затрат мощности, обладают определенным быстродействием.) Эта схема имеет максимальную длину пути распространения сигнала по девяти логическим уровням. Пороговое логическое устройство, изображенное на рис. 5.8, требует восемнадцати пороговых логических элементов и использует только 5 логических уровней. Также представляет интерес сравнение числа соединений, необходимых для данных схем. В случае булевой логической схемы эта величина составляет 116, против 70 для пороговых логических устройств. Это сравнение можно кратко подытожить тем утверждением, что пороговое логическое устройство имеет превосходство приблизительно в 2 раза в отношении числа логических уровней, числа логических элементов и числа соединений.

Подобным же образом можно спроектировать схему 8-разрядного умножителя-сумматора, умножающего два 8-разрядных числа, прибавляющего 21-разрядное число и выдающего 21-разрядный результат. Здесь коэффициенты объединения по входу и разветвления по выходу были увеличены до 8. Результаты, полученные для данной схемы (с данными по 2-разрядному умножителю-сумматору, показанному в скобках), можно увидеть на рис. 5.9 [29, 30]. Число вентилях в последнем случае по отношению к предыдущему увеличилось почти как 3:1, тогда как число логических уровней и соединений осталось на уровне 2:1. Указанная схема является в достаточной степени

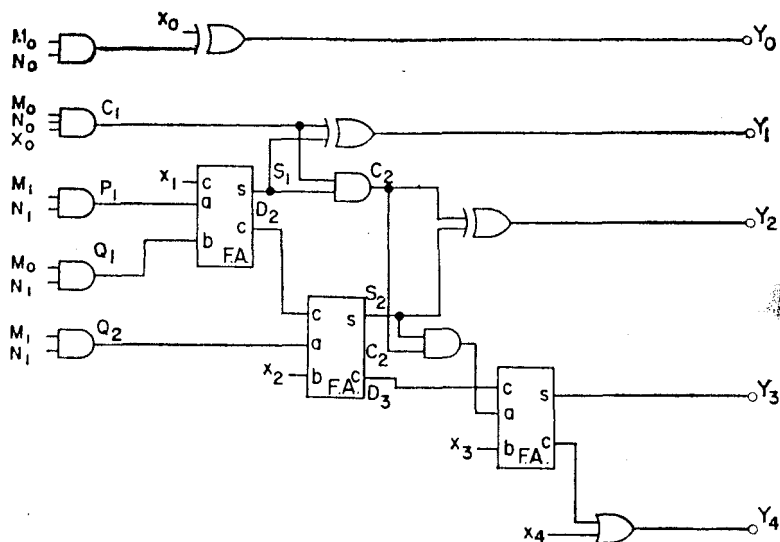


Рис. 5.7. Традиционный логический булев умножитель-сумматор с числом разрядов $(2 \times 2 + 5 \rightarrow 5)$. Полные сумматоры (F.A.) включают в себя два логических уровня и семь вентилях.

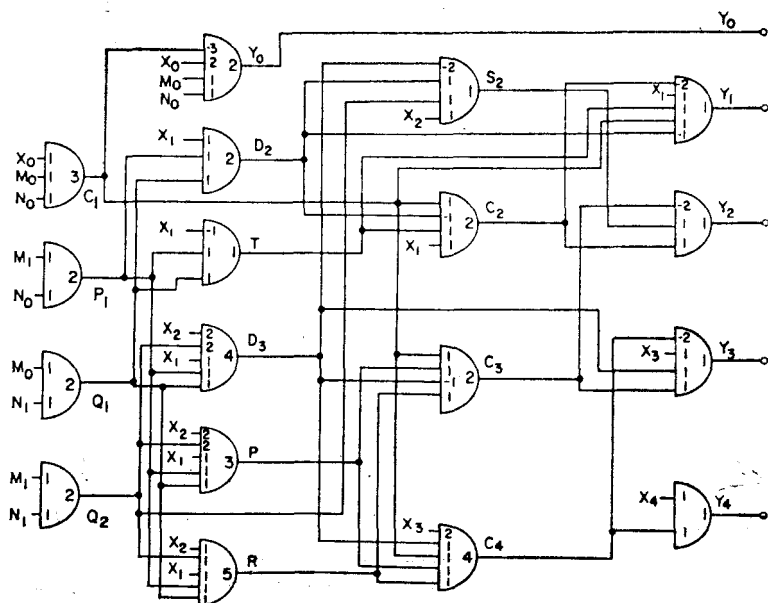


Рис. 5.8. Пороговый логический умножитель-сумматор с числом разрядов $(2 \times 2 + 5 \rightarrow 5)$.

сложной и возникает искушение сделать заключение, что результаты являются показателем того, чего можно достичь при реализации более сложных схем. Если рассматривать отношения скорости обработки к потребляемой мощности, то можно утверждать, что результаты также указывают на значительно большие преимущества указанных систем по сравнению с отдельным уровнем, элементом или набором каких-либо соединений, так как каждое из соответствующих отношений либо увеличивает скорость обработки, либо уменьшает потребление мощности, либо делает и то и другое.

Как указано выше, ожидается, что чисто оптические элементы и внутренние соединения будут необходимы в системах с внутренним пороговым кодированием, которые реализуют свои очевидные потенциальные преимущества по сравнению с возможностями чисто электронных устройств. А чтобы реализовать потенциальные преимущества в отношении геометрических размеров, вероятно, потребуется использовать методики интегральной или «почти интегральной» [31] оптики. Из числа целого ряда материалов, исследованных для указанной технологии (LiNbO_3 , стекло $/\text{SiO}_2/\text{Si}$ и т. д.), системы $\text{GaAs}/\text{GaAlAs}$ имеют наилучшие возможности для полной интеграции оптических источников, пороговых устройств и детекторов на одной подложке. Проблема состоит в том, что нелинейности материалов, необходимых для пороговых режимов работ в чисто оптических устройствах, по величине на два — четыре порядка должны превышать характеристики существующих однородных электрооптических материалов для систем, где оптический входной сигнал создается полупроводниковым лазером [32, 33]. Однако квантоворазмерные структуры на $\text{GaAs}/\text{AlGaAs}$ могут иметь необходимые нелинейные свойства при комнатной температуре, с приемлемыми скоростями переключения и энергетическими характеристиками, в том числе будучи включенными в состав плотно упакованных матриц ин-

	Традиционная логика		Пороговая логика	
Логических уровней	15	(9)	9	(5)
Логических элементов	805	(38)	289	(18)
Межэлементных соединений	2141	(116)	1175	(70)

Рис. 5.9. Сравнительные характеристики традиционной и пороговой логических схем умножителя-сумматора с числом разрядов $(8+8+21 \rightarrow 21)$. В скобках приведены данные для умножителя-сумматора с числом разрядов $(2 \times 2+5 \rightarrow 5)$.

тегрально-оптических структур на GaAs [34]. В настоящее время также проводятся исследования и других многообещающих материалов, включая системы на основе InSb [35, 36] и многослойные структуры на основе органических пленок Ленгмюра — Блоджетт [37].

5.3.2. Пример J-K-триггера

J-K-триггер является основным элементом традиционных последовательностных логических архитектур, которые запускаются либо синхронно от внешнего источника, либо работают в режиме внутренней синхронизации. Триггер имеет два входа J и K и выход Q; если оба J и K входных сигнала равны 0, то значение Q сохраняется; если только $J=0$, то Q равняется 0; если только $K=0$, то $Q=1$; если оба сигнала J и K равны 1, то Q принимает значение, противоположное тому, которое имелось ранее. На рис. 5.10 показано, как J-K-триггер можно было бы реализовать в устройстве, где фемтосекундные импульсы света сохраняют предыдущие значения выходного сигнала в оптической цепи обратной связи [18]. Следует отметить, что необходимо иметь максимальное число оптических путей, равное 6, и требуются 3 чисто оптических пороговых устройства; также показаны соответствующие весовые коэффициенты и пороги для операций J-K-триггера.

Архитектура с фемтосекундными ультракороткими импульсами [20, 38, 39] эквивалентна схеме с пороговыми логическими элементами, представленной на рис. 5.8, но обладает обратной связью. Так как данная архитектура входит в число наиболее общих и мощных разработок, ее характеристики и возможности обсуждаются далее более подробно.

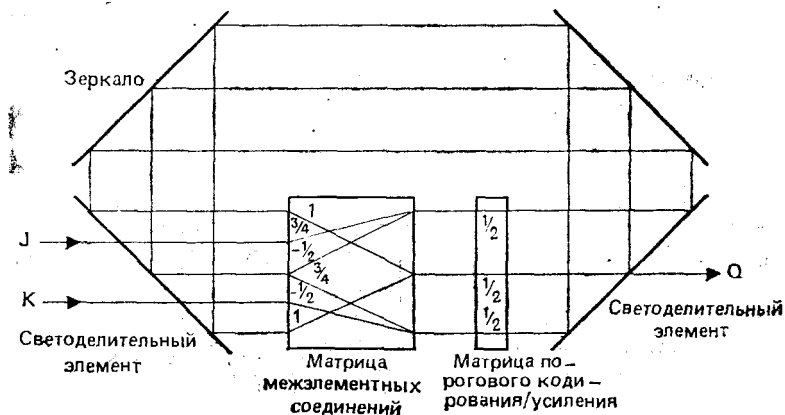


Рис. 5.10. Схема внутреннего порогового кодирования для J-K-триггера. Весовые коэффициенты показаны на матрице межсоединений; пороги указаны на матрице порогового кодирования/усиления.

5.4. Заключение

Архитектура с фемтосекундными ультракороткими импульсами, для которой на рис. 5.11 дана принципиальная схема, является подходящим объектом для изложения заключительных замечаний, связанных с пороговым кодированием и взвешиванием в оптических вычислениях. В данной разработке входные данные, включающие информацию об управлении и программировании, закодированы во входном пучке с помощью пространственной и временной модуляции. Оптическая матрица соединений (содержащая голограммы, линзы и т. д.) выполняет операции взвешивания, а матрица нелинейных пороговых устройств (как правило, выполняющая усиление сигнала) осуществляет операции порогового кодирования. Оптическая длина пути в цепи обратной связи превращает все устройства в последовательную вычислительную систему, в которой модулированные по координатам и времени фемтосекундные импульсы света могут циркулировать как «на конвейере». Синхронизация осуществляется либо асинхронно, с тактовой частотой, задаваемой временем пробега в петле обратной связи, либо синхронно, используя внешние тактовые сигналы. При сравнительно низких частотах электрические входные сигналы, подаваемые в нелинейное матричное устройство или в матрицу соединений (в последнем случае, возможно, через электрооптически управляемую решетку), могут подаваться в дополнение к оптическим входным, управляющим и программирующим данным.

Как показано в разд. 5.3.2, архитектура фемтосекундных импульсов эквивалентна схеме пороговых элементов с обратной связью. В некоторых вопросах она эквивалентна схемам, в которых соединения для чисто электронных логических вен-

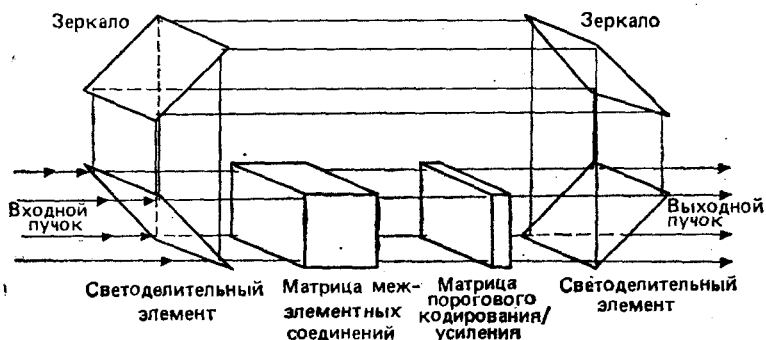


Рис. 5.11. Общий вид архитектуры устройства для оптических вычислений, в котором используются фемтосекундные ультракороткие импульсы света и внутреннее пороговое кодирование.

телей осуществляются оптически, например, в случае оптической междучиповой связи в СБИС [40]. В заключение заметим, что приведенная архитектура исследовалась в связи с нейронными системами [41, 42], которые, как упомянуто в 5.1.1, могут иметь исключительно «интеллектуальные» и гибкие рабочие характеристики. Для любой из этих интерпретаций (и обычно для архитектур, использующих матрицы логических устройств) рассеяние мощности является фактором, ограничивающим выбор рабочих режимов. Например, при максимальной рассеиваемой мощности в 10 Вт/см^2 , минимальной площади устройства в 10 мкм^2 и минимальной энергии переключения на единицу площади в 1 фДж/мкм^2 (что соответствует обоснованно ожидаемым характеристикам современной технологии GaAs [32]) удастся получить около 10^{15} логических операций в секунду на квадратный сантиметр (10^{15} переключений $\cdot \text{Гц/см}^2$), что по величине на два порядка больше, чем намечается получить согласно нынешним планам работ по СБИС. Однако если энергопотребление не представляет столь серьезной проблемы, что может быть либо в случае работы в режимах с малой нагрузкой, либо во «взрывном» режиме, то при минимальных переключающих временах в 10 пс возможно получить приблизительно 10^{18} переключений $\cdot \text{Гц/см}^2$.

В данной главе обсуждались возможности оптических вычислений в рамках представлений о взвешивании и пороговом кодировании, и приведены простые примеры схем с внешним и внутренним пороговым кодированием. Основная особенность оптических методов заключается в том, что они имеют превосходство в реализации соединений или операций взвешивания, но могут не обладать существенными преимуществами над чисто электронными методами при принятии решения или операциях порогового кодирования. Задача исследователей, таким образом, состоит в определении таких архитектур оптических вычислительных устройств, которые позволяют в наибольшей степени использовать возможности систем межэлементных соединений и реализуют преимущества оптических методов.

Часть исследований, рассмотренных в данной главе, получила финансовую поддержку программой СОИ, Агентством по инновации науки и технологии согласно контракту с управлением исследованиями ВМС № 00014-85-K-0479.

МНОГОЗНАЧНАЯ ПОРОГОВАЯ ЛОГИКА

*К. Морага, Отделение вычислительной техники,
Университет г. Дортмунда, Дортмунд, ФРГ*

6.1. Введение

Данная книга посвящена оптическим вычислениям, в то время как все предшествующие работы автора главы, касающиеся пороговой логики, были связаны с чисто электронными схемами. Тем не менее, находясь под сильным впечатлением от перспектив, открываемых оптическими вычислениями, он принял предложение и написал эту главу. По этой причине она объединяет общеобразовательные аспекты многозначной пороговой логики с наивными попытками разработать архитектуру, включающую в себя электрооптические устройства. Автор надеется, что читатели, целенаправленно занимающиеся проблемой оптических вычислений, найдут повод для дальнейших исследований теоретических вопросов, содержащихся в этой главе, и, более того, терпимо и с пониманием отнесутся к «наброскам» автора в области электрооптики.

Разработки в сфере оптических вычислений производят очень сильное впечатление, особенно с точки зрения предоставляемых ими особых возможностей для выполнения параллельной обработки с высокой скоростью, аналогового умножения, свертки, операций над матрицами и преобразования Фурье [1, 2, 3]. Однако довольно парадоксальной выглядит проблема обеспечения простой реализации в оптике функционально полного набора логических связей [4]. Тем не менее развитие электрооптических методов модуляции интенсивности света подготовило путь для появления двоичной пороговой логики [5, 6]. Известно, что двоичная пороговая логика является функционально полной и имеет дополнительную привлекательную черту — программируемость: изменение весовых коэффициентов может осуществляться в реальном времени для того, чтобы изменить передаточную функцию порогового устройства.

Посредством механизмов модуляции интенсивности может быть эффективно и точно выполнено аналоговое умножение, реализуемое с помощью электрических входных и оптических выходных сигналов. В этом случае умножение можно интер-

претировать как операцию в заданном интервале непрерывных значений логических переменных. Это обстоятельство и убеждает в пользу изучения электрооптической многозначной пороговой логики, где по сравнению со случаем непрерывных значений переменных не стоит так остро проблема идентификации «соседних» логических уровней. Такой подход может открыть новые альтернативы для развития оптических вычислений.

Соответственно в разд. 6.2 излагаются теоретические основы многозначной пороговой логики. Этот параграф завершается критическими оценками ее преимуществ и ограничений.

В разд. 6.3 описаны мультилинейно-разделяющиеся функции и предлагается архитектура, ориентирующаяся на реализацию с помощью электрооптических устройств.

Глава завершается обсуждением результатов, включающим предложения по дальнейшим исследованиям.

6.2. Многозначная пороговая логика

6.2.1. Теоретические основы

Исследования многозначной пороговой логики возникли в начальный период исследований пороговой логики. За важной работой [7] по описанию свойств пороговых (двоичных) функций последовали пионерские работы [8, 9] по троичной пороговой логике. На протяжении последующего десятилетия основное внимание исследователей привлекла именно троичная пороговая логика, что определялось, вероятно, появлением возможности для ее приборной реализации на основе дискретных полупроводниковых компонент. В числе наиболее важных результатов данного периода можно упомянуть работы, посвященные описанию свойств троичных пороговых функций [10, 11], подсчету и классификации всех трехместных троичных пороговых функций [12, 13], откуда следует, что существует 85 629 таких функций (в то время как рядом авторов независимо указывалось на существование 471 двухместных троичных пороговых функций), а также табличный метод реализации троичных пороговых функций с числом переменных, достигающим трех [13, 14].

Первыми публикациями по многозначной пороговой логике (за исключением троичной), по-видимому, являются работы [15—17]. Интересно заметить, что в [15] набор p корней единицы был использован в качестве области определения p -значных функций, а для исследования пороговых функций использовались методы гармонического анализа функций. Данный подход, однако, не получил дальнейшего развития в последующей литературе. С другой стороны, в [16] в качестве области

определения p -значной функции были использованы первые p неотрицательные корни, а также были введены понятия характеристического вектора многозначной пороговой функции и даны необходимые и достаточные условия того, что p -значная функция являлась пороговой функцией.

Следующее определение представляет собой упрощенный вариант материала, изложенного в работе [17].

Определение 6.1

Пусть $V = \{0, 1, \dots, p-1\}$. Функция $f: V^n \rightarrow V$ является n -местной p -значной пороговой функцией, если только существуют действительный вектор $W = (w_1, w_2, \dots, w_n)$, называемый весовым вектором, и действительный вектор $T = (t_1, t_2, \dots, t_{p-1})$, называемый пороговым вектором, такие что

$$\begin{aligned} W \cdot X \geq t_{p-1} &\Leftrightarrow f(X) = p-1, \\ t_{p-1} > W \cdot X \geq t_{p-2} &\Leftrightarrow f(X) = p-2, \\ &\vdots \\ t_1 > W \cdot X &\Leftrightarrow f(X) = 0, \end{aligned} \quad (6.1)$$

где $X = (x_1, x_2, \dots, x_n)$ и $W \cdot X$ обозначает внутреннее произведение W и X .

Становится ясно, что, если система (6.1) имеет решение, тогда она имеет бесконечно много решений. Большая их часть удовлетворяет более жесткой системе неравенств со знаком «>» вместо « \geq ».

Следует отметить, что определение не требует, чтобы функция устанавливала фактически многооднозначное соответствие. Если f не принимает заданного значения k в V , это только означает, что в V^n существуют X , такие что $f(X) = k$, и в (6.1) выражение $t_{k+1} > W \cdot X \geq t_k \Leftrightarrow f(X) = k$ все еще является формально правильным, но не имеет решений. (Функция $f: A \rightarrow B$ устанавливает нестрогое многооднозначное соответствие¹⁾, если для любого $b \in B$ существует по крайней мере одно $a \in A$, такое что $f(a) = b$.)

¹⁾ Согласно словарю по логике — Кондаков Н. Н. Логический словарь-справочник. — М.: Наука, 1975 — многооднозначное соответствие в строгом смысле подразумевает, что каждому элементу A соответствует только один элемент B , но каждому из элементов B соответствует более одного элемента из A . Следует заметить, что в данном случае автор гл. 6 дает интерпретацию более строгого понятия сюръективной функции, определение которой приведено, например, в [Кук Д., Бейз Г. Компьютерная математика: Пер. с англ. — М.: Наука, 1990, гл. 3]. — Прим. перев.

Определение 6.2

Пусть $f^{-1}(v) = \{X | f(X) = v\}$, $v \in V$. Тогда f является пороговой функцией, если только существует набор n -мерных параллельных гиперплоскостей в V^n , разделяющих $f^{-1}(1) \dots$ от $f^{-1}(p-1)$.

Рассмотрим пример трюичной пороговой функции на рис. 6.1.

Введем следующее обозначение. Если f — пороговая функция с весовым вектором W и пороговым вектором T , то запишем $f: (W; T)$ и назовем $(W; T)$ структурой f . Отсюда для примера на рис. 6.1 получим, что $f: ((1, 0, 3); (0,75, 2,75, 3,75))$.

Лемма 6.1

Пусть f является пороговой функцией со структурой $(W; T)$. Тогда $f: (kW; kT)$ справедливо в случаях, где масштабный множитель k является целым положительным числом.

Доказательство: Следует непосредственно из уравнения (6.1).

Лемма 6.2

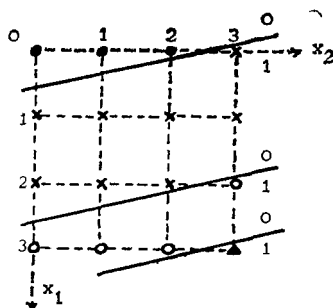
Для всех $x \in V$ положим $x = p-1-x$. Более того, определим $\alpha_i: V^n \rightarrow V^n$ с $\alpha_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{in})$, $1 \leq i \leq n$, $\beta_{ij}(X) = x_j \forall j \neq i$ и $\beta_{ii}(X) = \bar{x}_i$. Пусть $f: V^n \rightarrow V$ является пороговой функцией со структурой $(W; T)$. Функция от функции f по любому α_i также является пороговой функцией и имеет структуру $(W^{(i)}; T^{(i)})$, где $w_i^{(i)} = -w_i$, $w_k^{(i)} = w_k$ и $t_j^{(i)} = t_j - (p-1)w_i$, $1 \leq k \leq n$, $1 \leq j < p$, $k \neq i$.

Доказательство: Пусть $\delta(W, X) = \sum_{i=1}^n w_i x_i$. Тогда $W \cdot X = \delta(W, X) + w_i x_i$. Пусть $W^{(i)}$ будет таким, что $w_i^{(i)} = -w_i$ и $w_k^{(i)} = w_k$ для всех $k \neq i$. Отсюда следует, что

$$\begin{aligned} W^{(i)} \cdot \alpha_i(X) &= \delta(W, X) + w_i^{(i)} \cdot \bar{x}_i = \delta(W, X) - w_i(p-1-x_i) \\ &= \delta(W, X) + w_i x_i - (p-1)w_i = W \cdot X - (p-1)w_i. \end{aligned}$$

$f(x):$

	x_2			
x_1	0	1	2	3
0	0	0	0	1
1	1	1	1	1
2	1	1	1	2
3	2	2	2	3



$$X = (x_1, x_2)$$

$$W = (1, 0, 3)$$

$$T = (0,75, 2,75, 3,75)$$

$$\bullet := 0 \quad \times := 1 \quad \circ := 2 \quad \blacktriangle := 3$$

Рис. 6.1. Пример двуместной четырехуровневой пороговой функции.

Отсюда получаем

$$W \cdot X = W^{(i)} \cdot \alpha_i(X) + (p-1) \omega_i.$$

Положим в данном случае $t_p = \infty$ и $t_0 = -\infty$. Так как $f: (W; T)$, то можно записать

$$t_j > W \cdot X \geq t_{j-1} \Leftrightarrow f(X) = j-1, \quad j = 1, 2, \dots, p,$$

но

$$\begin{aligned} t_j > W \cdot X \geq t_{j-1} &\Leftrightarrow t_j > W^{(i)} \cdot \alpha_i(X) + (p-1) \omega_i \geq t_{j-1} \\ &\Leftrightarrow t_j - (p-1) \omega_i > W^{(i)} \cdot \alpha_i(X) \geq t_{j-1} - (p-1) \omega_i. \end{aligned}$$

Определим $t_j^{(i)} = t_j - (p-1) \omega_i$, $j = 0, 1, \dots, p$. Тогда

$$t_j > W \cdot X \geq t_{j-1} \Leftrightarrow t_j^{(i)} > W^{(i)} \cdot \alpha_i(X) \geq t_{j-1}^{(i)}.$$

Отсюда следует

$$\begin{aligned} t_j^{(i)} > W^{(i)} \cdot \alpha_i(X) \geq t_{j-1}^{(i)} &\Leftrightarrow f(X) = j-1 \\ &\Leftrightarrow f(\alpha_i(X)) = j-1, \end{aligned}$$

т. е. $f \alpha_i: (W^{(i)}; T^{(i)})$.

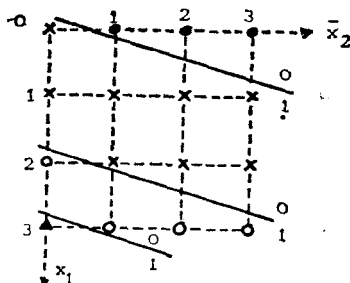
На рис. 6.2 показана функция от функции $f(\alpha_2(X))$, где f является функцией, ранее изображенной на рис. 6.1.

Следствие 6.1

Пусть $f: V^n \rightarrow V$ является пороговой функцией со структурой $(W; T)$. Более того, предположим, что некоторые ω_i имеют отрицательные значения, например $\omega_i = -u$, где $u > 0$. Тогда f реализуется в виде пороговой функции с положительными весовыми коэффициентами, которую вычисляют, находя дополне-

$f(\alpha_2(x))$:

	\bar{x}_2			
	0	1	2	3
0	1	0	0	0
1	1	1	1	1
2	2	1	1	1
3	3	2	2	2



$$\alpha_2(X) = (x_1, \bar{x}_2)$$

$$W^{(2)} = (1, -0.3)$$

$$T^{(2)} = (-0.15, 1.85, 2.85)$$

Рис. 6.2. Четырехуровневая пороговая функция, полученная из функции, заданной на рис. 6.1 путем замены второго аргумента на его дополнение.

ния к соответствующим аргументам x_i и увеличив все пороги на величину $(p-1)u$.

Лемма 6.3

Пусть $f: V^n \rightarrow V$ является пороговой функцией со структурой $(W; T)$. Тогда \bar{f} также является пороговой функцией и имеет структуру (W'', T'') , где $W'' = -W = (-\omega_1, -\omega_2, \dots, -\omega_n)$ и $T'' = (-t_{p-1}, \dots, -t_2, -t_1)$.

Доказательство:

$$t_j > W \cdot X \geq t_{j-1} \Leftrightarrow f(X) = j-1, \\ \Leftrightarrow \bar{f}(X) = (p-1) - (j-1) = p-j.$$

Умножая на (-1) , получим

$$-t_{j-1} > -W \cdot X = -t_j \Leftrightarrow -f(X) = -(j-1).$$

Пусть $W'' = -W$ и $t_i'' = -t_{p-i}$, $1 \leq i \leq p$. Более того, прибавим $p-1$ к обеим частям второго члена приведенного выше выражения. Это приводит к

$$t''_{p-j+1} > W'' \cdot X \geq t''_{p-j} \Leftrightarrow (p-1) - f(X) = (p-1) - (j-1) \\ \Leftrightarrow \bar{f}(X) = \overline{(j-1)} = p-j,$$

т. е. $\bar{f}: (W''; T'')$.

Пример

Пусть f — функция, представленная на рис. 6.1. Тогда справедливо следующее:

$$\bar{f}: ((-1, -0,3); (-3,75; -2,75; -0,75)).$$

Следствие 6.2

Пусть $f: V^n \rightarrow V$ является пороговой функцией со структурой $(W; T)$. Двойственная функция, заданная как $f^d = f \alpha_1 \alpha_2 \dots \alpha_n$, также является пороговой функцией и имеет структуру $(W; T^d)$, где для всех j , таких, что $1 \leq j < p$, справедливо равенство

$$t^d_j = (p-1) \sum_{i=1}^n \omega_i - t_{p-j}.$$

Доказательство

Следует непосредственно из лемм 6.2 и 6.3. Пример представлен на рис. 6.3.

Лемма 6.4

Пусть $f: V^n \rightarrow V$ является пороговой функцией. Ее структура $(W; T)$ при этом удовлетворяет строгому варианту уравнения (6.1). Пусть $L_i: V \rightarrow V$, $i=0, 1, 2, \dots, p-2$, так что $L_i(v) = i$, если $v = i+1$ и $L_i(v) = v$ для всех v , принадлежащих V . Тогда $L_i f: (W; T')$ справедливо для $t'_j = t_j$, если $j \neq i+1$ и

$t'_{i+1} = t_{i+2} - e$ (где e — положительная бесконечно малая величина).

Доказательство: Из определения L_i и строгого неравенства для f становится ясно, что

$$(Lif)^{-1}(i+1) = \emptyset,$$

$$(Lif)^{-1}(i) = f^{-1}(i) \cup f^{-1}(i+1)$$

$$\text{и } (Lif)^{-1}(v) = f^{-1}(v) \forall v \in V \setminus \{i, i+1\}.$$

Другими словами:

Не существует X такого, что $Lif(X) = i+1$.

Для всех X , таких что $f(X) = i$ или $f(X) = i+1$, $Lif(X) = i$.

Иначе $Lif(X) = f(X)$.

Если заменить элементы T на соответствующие элементы T' и $f(X)$ на $Lif(X)$ в уравнении (6.1), то изменяется все выражения, кроме двух, а именно

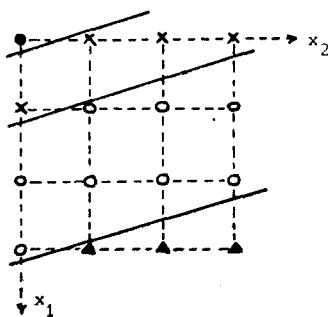
$$t'_{i+2} > W \cdot X \geq t'_{i+1} \Leftrightarrow Lif(X) = i+1,$$

$$t'_{i+1} > W \cdot X \geq t'_i \Leftrightarrow Lif(X) = i.$$

Первое из них не имеет решений, так как по определению не существует X , таких что $Lif(X) = i+1$. Левая сторона второго выражения соответствует $t_{i+2} - e > W \cdot X \geq t_i$ и включает в себя все X , такие что $f(X) = i+1$ или $f(X) = i$; но таковыми являются в точности все X , такие что $Lif(X) = i$. Отсюда следует, что второе выражение является справедливым, что завершает доказательство.

$f^d(X)$:

		x_2			
		0	1	2	3
x_1	0	0	1	1	1
	1	1	2	2	2
	2	2	2	2	2
	3	2	3	3	3



$$X = (x_1, x_2)$$

$$W = (1, 0.3)$$

$$T^d = (0.15, 1.15, 3.15)$$

Рис. 6.3. Двойственная функция, полученная из функции, показанной на рис. 6.1, и ее пороговая структура.

Следствие 6.3

Пусть f является таким, как указано в лемме 6.4. Пусть $U_i: V \rightarrow V$, $i=1, 2, \dots, p-1$ обладает тем свойством, что $U_i(v) = i$, если $v=i-1$ и $U_i(v)=v$ для всех других v , принадлежащих V . Тогда $U_i f: (W; T'')$ во всех случаях, где $t_j''=t_j$, если $j \neq i$ и $t_i''=t_{i-1}+e$.

Лемма 6.5

Пусть $f: (W; T)$ является p -значной пороговой функцией. При этом пусть $h: V^n \rightarrow V^n$ является взаимно однозначным¹⁾ соответствием. Тогда $f(h(X))$ также является пороговой функцией со структурой $(W^h; T)$, где W^h является весовым вектором, полученным путем перестановки весовых коэффициентов в W таким же способом, как элементы X переставляются под h .

Доказательство: Следует непосредственно из уравнения (1) и свойства коммутативности суммы в $W \cdot X$.

В качестве заключительного замечания к данному теоретическому основанию следует напомнить, что авторы [18] доказали функциональную полноту многозначной пороговой логики. Это означает, что любая многозначная система может быть реализована только с помощью пороговых функций. Однако это не означает, что многозначные системы следует реализовывать именно с помощью пороговых функций.

6.2.2. Критические оценки возможностей

Многозначная пороговая логика, как можно убедиться на примерах, приведенных на рис. 6.1—6.4, привлекательна тем, что она позволяет дать простое представление нетривиальных функций. Более того, функциональная полнота многозначной пороговой логики предоставляет возможность чисто пороговой реализации сложных многозначных систем.

С другой стороны, многозначная пороговая логика выглядит достаточно непривлекательно с позиции «комбинаторного взрыва» многозначных функций. Известно, что в двоичном случае существуют 16 двухместных функций, из которых 14 являются пороговыми. В троичном случае, однако, имеется уже $3^9=19\,683$ двухместных функций и (только) 471 из них являются пороговыми [12]. В четвертичном случае имеется 4^{16} двухместных функций (около $4,3 \cdot 10^9$ функций), из которых только 18 184 являются пороговыми. При рассмотрении функции трех переменных 104 из них (около 40% являются пороговыми), в то время как из $7,6 \cdot 10^{12}$ троичных функций только 85 629 являются пороговыми [12]. В заключение отметим, что

¹⁾ В более строгом смысле — биективной (bijective) функцией (см. Кук Д., Бейз Г. Компьютерная математика: Пер. с англ. — М.: Наука, 1990, гл. 3), — *Прим. перев.*

$L_1 f(X):$

		x_2			
		0	1	2	3
x_1	0	0	0	0	1
	1	1	1	1	1
	2	1	1	1	1
	3	1	1	1	3

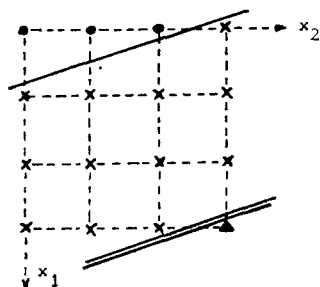
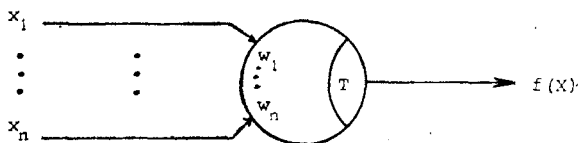
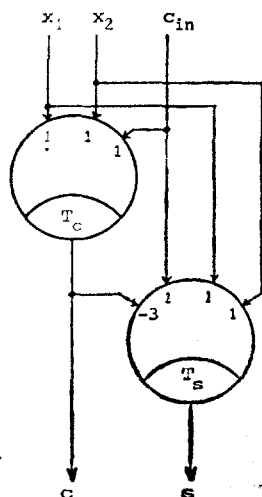
 $X = (x_1, x_2)$ $W = (1, 0.3)$ $T = (0.75, 3.74, 3.75) \quad \epsilon = 0.01$

Рис. 6.4. Пример сохранения пороговой разделимости при преобразовании области определения.

а:



б:



x_1	x_2	c_{in}	c	s	y_1	y_2
0	0	0	0	0	0	0
0	1	0	0	1	1	1
0	2	0	0	2	2	2
1	1	0	0	2	2	2
1	2	0	1	0	3	0
2	2	0	1	1	4	1
1	1	1	1	0	3	0
1	2	1	1	1	4	1
2	2	1	1	2	5	2

$$y_1 = x_1 + x_2 + c_{in}$$

$$y_2 = y_1 - 3c$$
 $T_c = (2.5, \infty)$ $T_s = (0.5, 1.5)$

Рис. 6.5. а — схематическое изображение многоуровневого порогового вентиля; б — схема построения трюичного полного сумматора.

относительное число многозначных пороговых функций является незначительным; тем не менее абсолютное число таких функций достаточно велико, чтобы служить стимулом для поисков подходящих классов задач, которые могут быть эффективно решены с помощью многозначной пороговой логики. Хотя все это до сих пор представляет недостаточно проработанный вопрос, уже сейчас имеются обнадеживающие результаты. Один из ярчайших примеров эффективного применения многозначной пороговой логики — это реализация p -значного полного сумматора, построенного на основе только двух пороговых вентилях и показанного на рис. 6.5 для $p=3$. (Для более подробного ознакомления с вопросом следует обратиться к работам [9, 20].)

6.2.3. Многозначная пороговая логика и электрооптические устройства

Рассмотрим теперь реализацию пороговых вентилях с помощью оптических или электрооптических устройств. Для непосредственной реализации пороговой функции необходимо осуществить операции перемножения сигналов и весовых коэффициентов, суммирования взвешенных сигналов и квантования взвешенных сумм по отношению к набору заданных пороговых значений (в процессе, названном «пороговым кодированием»).

Проблема умножения может быть решена традиционным образом в гибридных системах с помощью электрооптически моделируемых волноводов [3, 5, 21]. Данный процесс основан на управляемом эффекте поглощения оптического носителя информации. Очевидно, что это ограничивает возможность выполнения операции умножения лишь диапазоном величин от 0 до 1 (см. лемму 6.1). Однако масштаб весовых коэффициентов и пороговых функций может изменяться в любом желаемом диапазоне. (Наряду с этим может потребоваться изменить масштаб для «подгонки» динамического диапазона операции умножения к наиболее линейной части рабочей характеристики модулятора.)

Сложение оптических сигналов представляет простую проблему и может быть решено с помощью светоделительных элементов или линз.

Чтобы решить проблему порогового кодирования, в идеальном случае следовало бы стремиться к созданию оптического переключающего устройства с передаточной характеристикой вход/выход, имеющей вид «лестницы». Для этого подошло бы любое оптическое пороговое устройство с управляемым порогом. Однако до настоящего времени отсутствовали работоспособные оптические переключающие элементы како-

го-либо вида, хотя некоторые экспериментальные результаты, представленные в [4], вполне обнадеживают. Это означает, что пороговое кодирование все еще должно выполняться электронными способами. Такие методики хорошо известны. (Например, на рис. 6.6 показан I^2L -пороговый детектор¹⁾, предназначенный для реализации четырехуровневой логики.)

Хотя трехкомпонентные узлы порогового вентиля могут быть реализованы указанным способом, общий результат не кажется удовлетворительным, в особенности из-за слишком высокой степени «гибридности», обусловленной наличием электронного выходного канала. Далее будет обсуждаться расширенный вариант пороговой логики, предполагающий реализацию ее посредством разложения по функциям определенного вида. Это направление выглядит весьма многообещающим.

6.3. Мультилинейные разделяющиеся функции

В предыдущем разделе обсуждался вопрос о том, что многозначные пороговые функции обладают двумя недостатками: их сравнительно небольшим числом и трудностью осуществления оптического порогового кодирования. Ниже описывается класс мультилинейных разделяющихся функций, которые являются намного более широким видом функций и включают в себя пороговые функции. Более того, по самому определению этих функций предполагается разная архитектура при их реализации, и это, как будет видно, дает подходящую возможность ис-

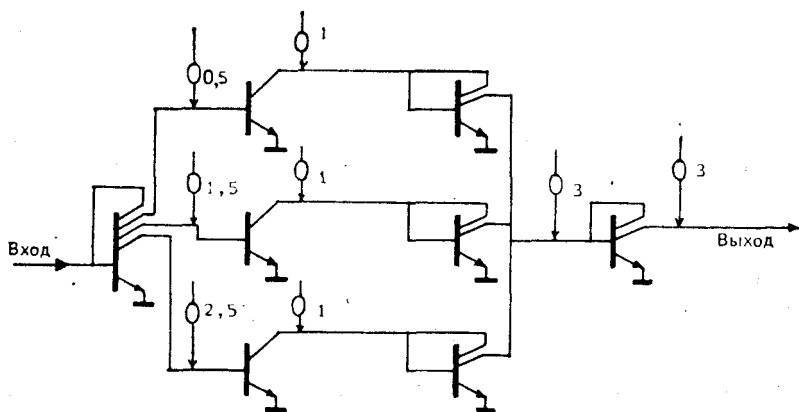


Рис. 6.6. Четверичный пороговый I^2L -детектор, используемый в качестве четырехуровневой восстанавливающей постоянной составляющей.

¹⁾ I^2L — устройство, реализуемое на интегральных инжекционных логических схемах. — Прим. перев.

пользовать уже существующие электрооптические методы и устройства.

Определение 6.3 [13]

Функция $g: V^n \rightarrow \{0, 1\}$, являющаяся при этом пороговой, называется линейной разделяющей (ЛР) функцией. Две функции g_1 и $g_2: V^n \rightarrow \{0, 1\}$, являющиеся ЛР-функциями с одинаковыми весовыми коэффициентами, называются изобарическими функциями (или изобарами).

Определение 6.4 [22]

Мультимножество представляет собой набор элементов, возможно включающий в себя повторения этих элементов (в отличие от множества, представляющего набор различных элементов). Количеством элементов мультимножества (или мощность мультимножества) является полное число элементов в наборе.

Определение 6.5 [23]

Функция $f: V^n \rightarrow V$ является мультилинейной разделяющейся (МР), если существует мультимножество $G = \langle g_i | g_i: V^n \rightarrow \{0, 1\}, \text{ где } g_i, \text{ являющееся ЛР-функцией и } 1 \leq i < p \rangle$, такое что

$$\sum_{i=1}^{p-1} g_i(X) = f(X). \quad (6.2)$$

МР-функция $f: V^n \rightarrow V$ является монотонной мультилинейной разделяющейся (ММР), если для всех величин X в V^n

$$\begin{aligned} \hat{f}(X) < i &\Leftrightarrow g_i(X) = 0 \\ \text{и } \hat{f}(X) \geq i &\Leftrightarrow g_i(X) = 1. \end{aligned} \quad (6.3)$$

Становится понятным, что, если все количество g_i вспомогательных функций для ММР-функции f являются изобарами, тогда f является пороговой функцией.

С точки зрения геометрии функция $f: V^n \rightarrow V$ является ММР, если существует множество *необязательно параллельных* n -мерных гиперплоскостей, отделяющих $f^{-1}(0)$ от $f^{-1}(1)$, ..., ..., $f^{-1}(p-1)$.

Авторы [19] показали, что в троичной логике существует 703 двухместных ММР-функций и 532 485 трехместных ММР-функций; в четверичной логике существует 61 160 двухместных ММР-функций. Даже если эти результаты оценок, рассматриваемых лишь как часть полного числа таких функций соответствующих аргументов, все еще не имеют особого смысла, то все равно они показывают определенное улучшение ситуации по сравнению с соответствующим числом пороговых функций. В то время как имеющиеся данные, несомненно, слишком огра-

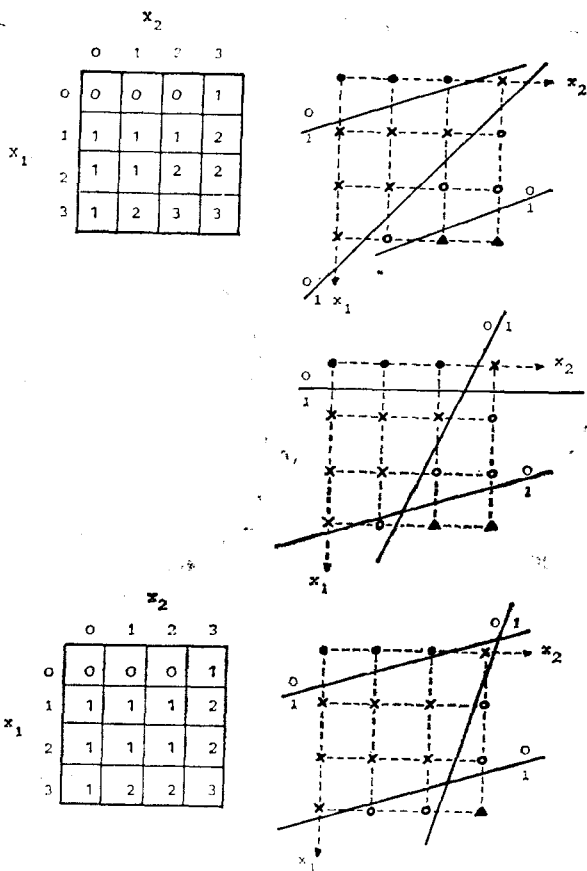


Рис. 6.7. а — монотонная мультилинейная разделяющаяся функция; б — вариант немонотонного мультилинейного разделения функции, показанной на рис. 6.7, а; в — мультилинейная разделяющаяся функция (немонотонная).

ничены для того, чтобы делать определенные выводы, все же можно предполагать, что по мере увеличения числа аргументов число ММР-функций возрастает намного быстрее, чем число пороговых функций.

Проверка (на немонотонность) мультилинейной разделимости может потребовать очень больших затрат времени из-за возможности существования большого числа различных способов разделения функций (см., например, рис. 6.7, а, б). С другой стороны, проверка монотонности мультилинейных разделяющихся функций является простой и быстрой процедурой и пригодна к (частично) параллельной обработке. Тогда это обеспечивает эффективный метод проверки пороговой разделимости p -значных функций [17, 19].

Лемма 6.6 [23]

Пусть $f: V^n \rightarrow V$ будет ММР-функцией. Функции, полученные перестановкой или дополнением аргументов в f , или дополнением f , также являются ММР-функциями.

Лемма 6.7

Пусть $f: V^n \rightarrow V$ является ММР-функцией, где $f(X) = \sum_{i=1}^{p-1} g_i(X)$.

Тогда \bar{f} также является ММР-функцией и

$$\bar{f}(X) = \sum_{i=1}^{p-1} g'_i(X),$$

где $g'_i(X) = 1 - g_i(X)$, $1 \leq i < p$, т.е. g'_i является двоичным дополнением к g_i .

Доказательство: Умножим первое уравнение на (-1) и прибавим $p-1$ к обеим частям.

На рис. 6.7 показаны примеры МР- и ММР-четвертичные функции.

Из определения 6.5 становится ясно, что реализация p -значных МР-функций (что включает пороговые функции) сводится к $p-1$ уровневым ЛР-функциям. Но ЛР-функции являются просто двоичными пороговыми функциями с p -значными аргументами. Кроме того, двоичное пороговое кодирование имеет адекватную (электро)оптическую реализацию [4, 5]. Учет этих факторов приводит к выработке общей архитектуры МР-функций, показанной на рис. 6.8.

В данной архитектуре используются входные каскады схемы, включающие в себя $n \cdot (p-1)$ двухканальных электрооптических волноводных модуляторов [5]. Любой x -сигнал должен быть подвергнут предварительной электронной обработке, чтобы получить требуемый коэффициент разветвления по выходу $(p-1)$ и правильное изменение масштаба, чтобы добиться совместимости с линейным динамическим диапазоном подсоединенных оптических устройств. Эти волноводы будут выдавать выходной световой сигнал с амплитудой, пропорциональной $x_i \omega_{j,i}$, $1 \leq i \leq n$, $1 \leq j \leq p$, которая будет изменяться при введении в систему $p-1$ линз. Выходной световой сигнал j -й линзы представляет собой $\sum_{i=1}^n x_i \omega_{j,i}$. Чтобы выполнить необходимое

(двоичное) пороговое кодирование, предлагается использовать нелинейные оптические устройства [4]. Дополнительный входной сигнал будет необходим здесь для смещения сигнала взвешенной суммы, обеспечивая компенсационную точную настройку порогов и получение чистых логических 1 для всех X , где значения соответствующих $g_i(X)$ являются истинными.

Нелинейные оптические устройства выдают на выход множество $p-1$ функций g_i , которые будут складываться

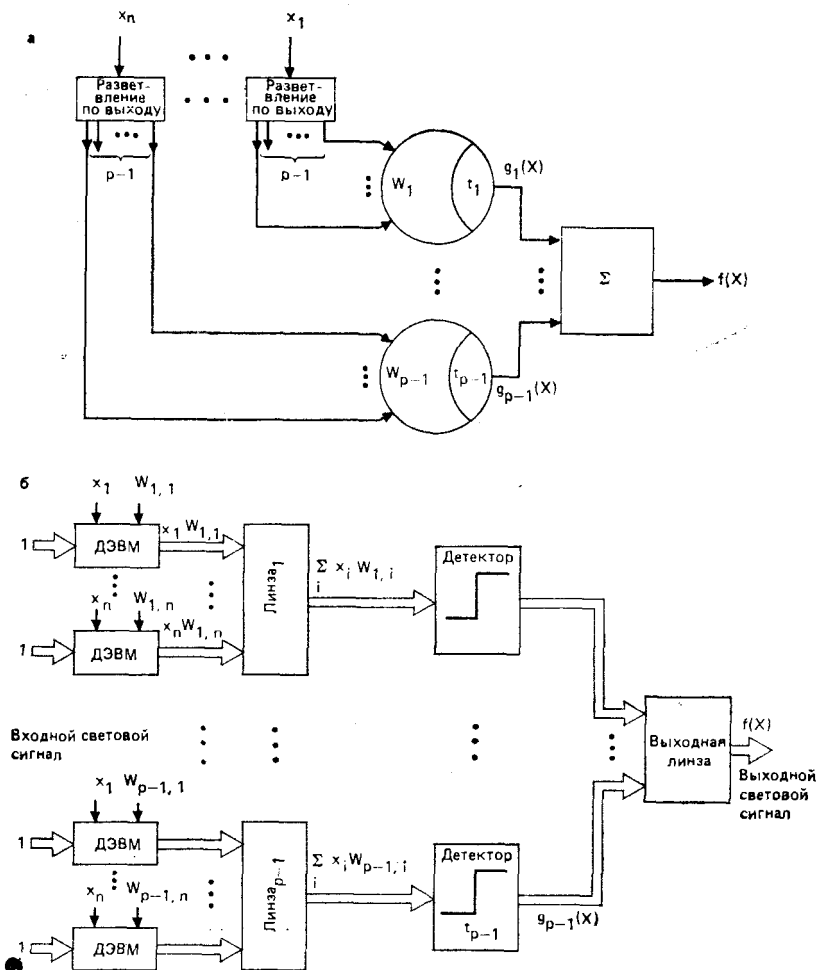


Рис. 6.8. а — общая модель построения r -значной МР-функции f в виде суммы $p-1$ ЛР-функций g_1, \dots, g_{p-1} ; б — общая схема реализации r -значных МР-функций в электрооптических устройствах. ДЭВМ — двухканальный электрооптический волноводный модулятор.

посредством последней из линз в схеме, чтобы получить интенсивность светового пучка, соответствующую $f(X)$.

Двоичное пороговое кодирование может также выполняться электронным способом с использованием, например, фотодиода. Сигналы, содержащиеся теперь в виде электрического сигнала и представляющие вспомогательные функции g_i , могут суммироваться для представления f в виде электрического сиг-

нала. Данный сигнал в свою очередь может быть использован для модуляции оконечного светового пучка с помощью другого волновода.

6.4. Перспективные разработки

Мультилинейные разделяющиеся функции фактически предлагают общий метод реализации многозначных функций путем разложения по двоичным вспомогательным ЛР-функциям. Далее эту идею можно развить, используя «обобщенное разделение» следующим способом:

Пусть $f: V^n \rightarrow V$. Определим мультимножество следующим образом: $G = \langle g_i | g_i: V^n \rightarrow \{0, 1\}, g_i \text{ является ЛР-функцией, } 1 \leq i \leq K \rangle$ при $K \geq p$. Определим $Q: V^n \rightarrow V^n$; $Q = (q_1, q_2, \dots, q_n)$, где $q_i: V \rightarrow V, 1 \leq i \leq n$. Если для всех X , принадлежащих V^n , выполняется

$$\bigcap_{i=1}^{K-1} g_i(Q(X)) = f(X),$$

тогда можно сказать, что f выражается с помощью обобщенного разделения, где Γ — замкнутая форма определенных парных сочетательных процедур. С геометрической точки зрения рассмотрим использование n -мерных (необязательно параллельных) гиперповерхностей вместо гиперплоскостей. Одна или более гиперповерхностей должны разделить некоторое множество $f^{-1}(v)$ от соседнего $f^{-1}(v+1)$ (или от $f^{-1}(v+1)$), так как полагаем $K \geq p$. Наконец, для любого X и i значения $g_i(X)$ будут составлены соответственно Γ .

Некоторые из этих обобщений исследовались в абстрактном виде или в контексте I^2L -технологии [23, 24, 25, 26]. Однако их приемлемость для электрооптических систем еще предстоит исследовать.

Для того чтобы проиллюстрировать некоторые аспекты обобщенного разделения, рассмотрим случай функции вычисления максимума. Без потерь общности для ниже изложенного используем троичную логику и нормализуем ее на интервале $[0, 1]$, определяя $V = \{0, 0.5, 1\}$. Становится ясно, что максимум не является МР-функцией. Однако, как показано на рис. 6.9, если допустить существование двух вспомогательных ЛР-функций (вместо одной), позволяющих разделить $f^{-1}(0,5)$ и $f^{-1}(1)$, то можно разложить $f(X)$ следующим образом:

$$f(X) = g_1(X) + g_2(X) + g_3(X),$$

где

$$g_i: V^2 \rightarrow \{0, 0,5\}, \quad i = 1, 2, 3,$$

$$g_1: (W_1; t_1) = ((1, 1); 0,25),$$

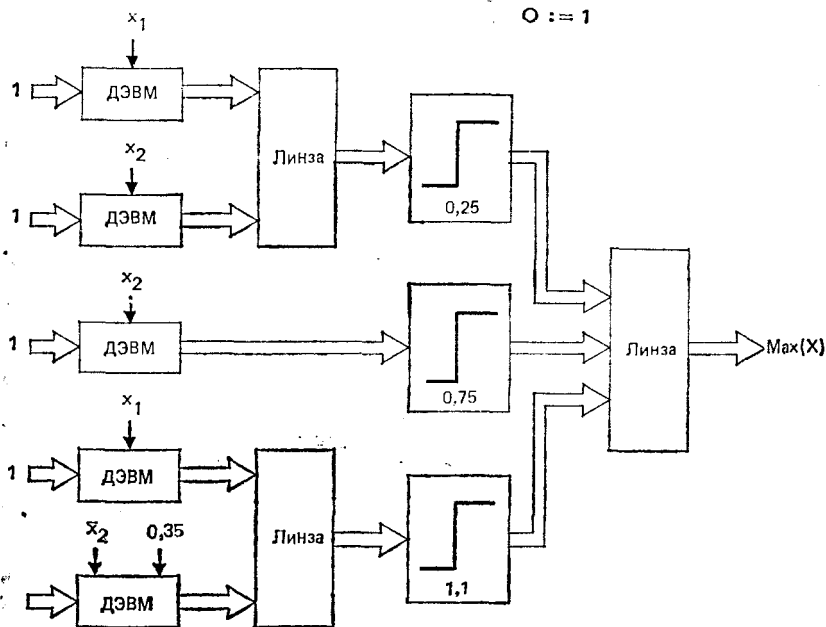
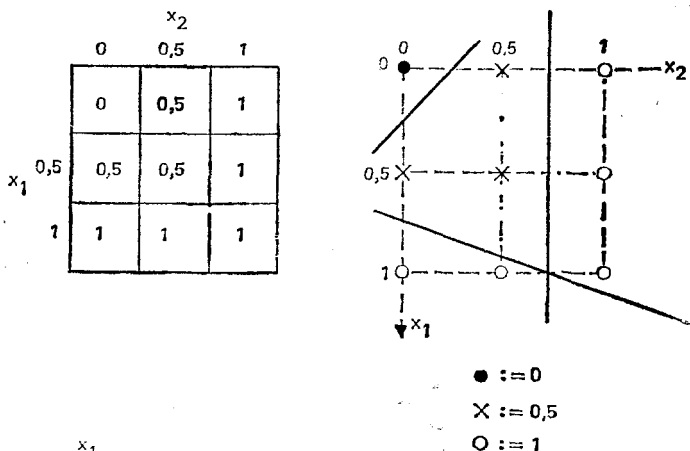


Рис. 6.9. Реализация функции определения максимума в четверичной логике с помощью обобщенного разделения.

$$g_2 : (W_2; t_2) = ((0, 1); 0,75),$$

$$g_3 : (W_3; t_3) = ((1, -0,35); 0,75).$$

Видно, что для реализации функции максимума можно использовать архитектуры, показанные на рис. 6.8, если при электронной предобработке получают два экземпляра x_2 и один экземпляр x_1 . Это следует из леммы 6.2:

$\alpha_2 g_3 : (W_3'; t_3') = ((1, +0.35); 1.1)$, а отрицательное значение весового коэффициента не является необходимым.

С другой стороны, если используются единичные ЛР-функции с диапазоном $\{0, 1\}$ для предобработки x_2 , то можно взять разделение, показанное на рис. 6.10, а.

Пусть $Q = (q_1, q_2)$, где q_1 — проекция x_1 (т. е. $q_1(X) = x_1$), и q_2 определяется следующим образом:

$q_1(X)$: если $x_2 \leq 0,5$, то равно 0, иначе — 1.

Тогда видно, что

$g_1(X) : ((1, 1); 0,25)$ $g_2(Q(X)) : ((1, 1); 0,75)$,

и для всех X , принадлежащих V^n ,

$f(X) = g_1(X) + g_2(Q(X))$.

Доказательство этого утверждения можно увидеть в табл. 6.1.

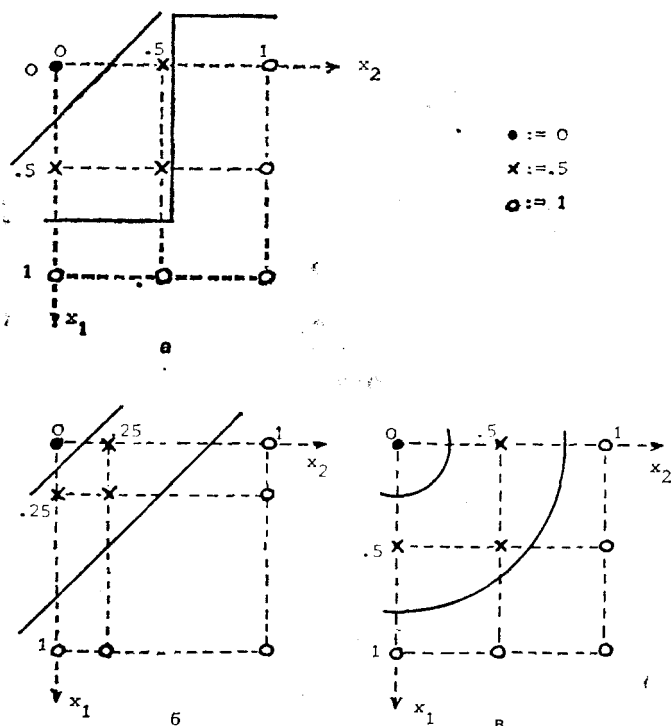


Рис. 6.10. а — обобщенное разделение функции определения максимума в четверичной логике с помощью предобработки аргумента; б — линейное разделение функции определения максимума после возведения в квадрат; в — квадратичное разделение функции определения максимума.

Таблица 6.1. Таблица истинности для разделения, представленного на рис. 6.10, а

x_1	x_2	$g_2(X)$	$g_1(X)$	$g_2(Q(X))$	$f(X)$
0	0	0	0	0	0
0	0,5	0	0,5	0	0,5
0	1	1	0,5	0,5	1
0,5	0	0	0,5	0	0,5
0,5	0,5	0	0,5	0	0,5
0,5	1	1	0,5	0,5	1
1	0	0	0,5	0,5	1
1	0,5	0	0,5	0,5	1
1	1	1	0,5	0,5	1

Наконец, если для преобразования электрического входного сигнала в световой используются нелинейные модуляторы, как это предполагалось ранее в архитектуре для МР-функций, то можно получить случай, показанный на рис. 6.10, б. Становится ясно, что функция определения максимума является «непосредственно порогово-реализуемой». В этом частном случае нелинейная модуляция может быть просто реализована путем возведения в квадрат. Входные каскады в этом случае должны состоять из трех каскадируемых двухканальных электрооптически модулируемых волноводов или брегговских дифракционных элементов, используемых для вычислений $x_i \cdot x_i \cdot w_i$. Предобработка электрических входных сигналов x_i тогда должна включать необходимый удвоенный коэффициент разветвления.

Это предложение является особым случаем полиномиальной разделимости [23, 25, 26] и недавно предложенного метода, где возможности оптического квадратичного порогового логического двоичного вентиля расширяются до уровня многозначной логики. На рис. 6.10, в можно явно увидеть квадратичное разделение функции вычисления максимума, где $y = x_1^2 + x_2^2$. Видно, что выполняется следующее:

$$y > 0,8 \Leftrightarrow f(X) = 1,$$

$$0,8 > y > 0,25 \Leftrightarrow f(X) = 0,5.$$

В заключение автору хочется выразить уверенность в том, что возможности многозначной пороговой логики и ее обобщения заслуживают дальнейшего исследования применительно к электрооптическим устройствам.

В области электрооптики возможности параллельной обработки и быстрого, точного вычисления с помощью аналогового умножения и сложения потенциально делают весьма привлекательной логику, реализованную на электрооптических уст-

ройствах. В данной главе показано, что многозначная пороговая логика и ее обобщения обладают некоторыми весьма интересными свойствами, но, по-видимому, могут непосредственно применяться лишь в довольно ограниченном числе областей. Вероятно, что для реализации сложных многозначных систем на основе электрооптических устройств могут потребоваться другие виды логических операций.

Автор весьма признателен проф. Р. Арратуну, убедившему его в целесообразности написания этой главы, за его терпеливое руководство, осуществляемое по мере углубления автора в волнующий мир электрооптики.

Автор благодарен Институту инженеров по электротехнике и радиоэлектронике за разрешение использовать в данной главе в переработанном виде некоторые результаты, ранее опубликованные в работе «Расширение многозначной пороговой логики», Материалы 9-го Международного симпозиума по многозначной логике, Бэт, Англия, 1979, ИИЭР.

В заключение, что не менее важно, автор благодарен редакторам за тщательно выполненную работу. Их предложения, несомненно, улучшили первоначальный вариант главы.

ЧАСТЬ III

СИСТОЛИЧЕСКИЕ ПРОЦЕССОРЫ И ОПТИЧЕСКИЕ ЛОГИЧЕСКИЕ МАТРИЦЫ

Глава 7

ОПТИЧЕСКОЕ УМНОЖЕНИЕ МАТРИЦ

*С. Кэртрайт¹⁾, Исследовательский институт
Университета г. Дайтона, Отделение прикладной
физики, Университет г. Дайтона, Дайтон, шт. Огайо*

7.1. Введение

Оптические вычисления, под которыми подразумевают выполняемые оптическими методами операции с дискретными числовыми данными, являются новинкой в долгой истории развития оптической обработки сигналов. Утверждения о том, что оптические методы могут успешно конкурировать и теоретически превзойти по своим возможностям электронные методы обработки данных, впервые привлекли серьезное внимание в середине 1970-х гг. [1, 2], а в последнее время в этом направлении возник настоящий шквал публикаций. Сначала может показаться, что электромагнитные поля оптического диапазона непригодны для реализации цифровой логики, так как они распространяются линейным и непрерывным образом, в то время как поток электронов в цепи может быть просто преобразован в дискретные двоичные уровни. Однако имеются три свойства оптики, которые делают ее привлекательной для цифровых вычислений. Первое — это широкая полоса частот оптических источников, которая может для полупроводниковых лазеров достигать гигагерц. Второе — это широкая полоса пространственных частот. Двумерная оптическая система может иметь крайне большое число элементов, разрешающих изображение, каждый из которых можно рассматривать как отдельный канал связи, а все они параллельно передают сигнал в одной и той же системе. В случае некогерентного освещения все разрешающие ячейки оптической системы являются взаимно независимыми. При освещении когерентным светом каналы являются связанными между собой, что приводит к исключительно высокой степени организации межэлементных соединений. Третьей, относящейся к оптическим соединениям, характеристикой является отсутствие интерференции при распространении сигналов, что иногда описывают как «возможность пересечения оптических проводов». Два оптических поля могут распространяться друг через друга, не оказывая взаимного влияния. Эти

¹⁾ В настоящее время работает в Исследовательском институте окружающей среды оз. Мичиган, Техническое бюро г. Дайтона, Дайтон, шт. Огайо.

характеристики дают возможность обрабатывать данные при больших скоростях и с недостижимым для электронных устройств уровнем организации связи между различными элементами.

Чтобы воспользоваться указанными преимуществами оптики, недостаточно заменить электронные компоненты на оптические. Фотон не является просто прозрачным электроном. Должны быть найдены специальные способы работы с данными, позволяющие использовать эти преимущества оптики и оптических элементов.

Обработка данных, как можно полагать, осуществляется на трех уровнях составных компонент компьютера. Простейший — это уровень вентиляей, в которых двоичные переключатели осуществляют булевы логические операции. Исследования на данном уровне концентрировались на разработке оптических бистабильных устройств, образующих вентили [3], и оптических методах образования соединений между оптическими или электронными логическими вентилями [4]. Наивысший уровень — это уровень процессора, на котором в едином узле выполняются целые алгоритмы. Традиционные оптические процессоры работают именно на данном уровне, выполняя, например, преобразование Фурье за один проход света через линзы. На уровне регистра осуществляется синергетическая обработка чисел и блоков чисел. Этот уровень организации обработки данных превосходит просто эксплуатацию набора вентиляей, но операции достаточно просты, и они могут быть сгруппированы так, чтобы образовать большое число операций более высокого уровня.

7.2. Цифровая оптика

7.2.1. Алгоритм цифрового умножения с помощью вычисления аналоговой свертки (ЦУАС)

Одна из целей цифровых оптических вычислений состоит в достижении большей гибкости системы, чем у их аналоговых предшественников. Особенность оптических компьютеров состоит в том, что они скорее выполняют не монолитные операции, а ряд простых операций, которые можно объединить для выполнения широкого круга задач. Однако в данном случае это не так плохо, поскольку при построении оптических процессоров, осуществляющих функции регистра, их возможности поднимутся на качественно новый уровень. С этой точки зрения матричное умножение (под которым подразумевают либо умножение матрицы на вектор, либо матрицы на матрицу), возможно, является наиболее полезной операцией среднего уровня из числа тех, которые только можно придумать. Многие сложные проблемы, например калмановское фильтрование,

можно представить в виде серии операций с матрицами. Матричное умножение представляет собой не что иное, как упорядоченные серии операций умножения и сложения, каждое из которых легко выполнить оптически. Далее обсуждение ограничено случаем некогерентной оптики, так что интерес будет представлять интенсивность света, а фазу рассматривать не потребуется. Используя термины из области оптики, числа представляют уровнями интенсивности света, или уровнями пропускания пространственных модуляторов света. Если свет с интенсивностью t_1 проходит через модулятор с коэффициентом пропускания t_2 , результирующая интенсивность света составляет $t_1 t_2$. Сложение осуществляется путем суммирования большого числа световых пучков на единственном фотодетекторе. В литературе описан ряд оптических матричных умножителей, использующих уровни интенсивности для представления целых чисел [5, 6]. Реализация всего этого затрудняется ограниченным динамическим диапазоном оптической обработки. На практике удастся получать, управлять и детектировать около 500 дискретных уровней. Это ограничивает точность вычислений примерно 8 разрядами. Точность в 8 бит является удовлетворительной лишь в редких случаях, что особенно проявляется в задачах с большим числом шагов, когда погрешности накапливаются. Как хорошо известно из области применения ЭВМ, решение должно быть представлено в цифровой форме. Каждое число представляют последовательностью цифр, каждая из которых имеет очень маленький динамический диапазон, а операции проводят над отдельными цифрами.

В работе [7] было показано, что умножение двух чисел эквивалентно выполнению свертки над их цифровыми представлениями в том случае, если результат операции свертки интерпретируется в смешанном формате, в котором для основания числа b отдельные цифры могут иметь величины, большие чем b . В общем, если возникает желание умножить по основанию b два N -разрядных числа, представленных в виде

$$x = \sum_{n=1}^N x_n b^{n-1}, \quad y = \sum_{m=1}^N y_m b^{m-1},$$

где x_n и y_m — отдельные цифры, то результат выглядит так:

$$\begin{aligned} xy &= \sum_{n=1}^N \sum_{m=1}^N x_n b^{n-1} y_m b^{m-1} = \sum_n \sum_{m=1}^N x_n y_m b^{(n+m-1)-1} = \\ &= \sum_{k=1}^{2N-1} \left(\sum_{n=1}^k x_n y_{k-n+1} \right) b^{k-1}, \end{aligned}$$

где $k = n + m - 1$. Теперь рассмотрим цифры в числах x и y , представляющие собой две выбранные функции, которые сле-

дует перемножить. Для выполнения свертки запишем цифры числа y в обратном порядке и разместим их после цифр числа x . Первым значением свертки является x_1y_1 , вторым $x_1y_2 + x_2y_1$ и т. д. до x_Ny_N ; k -е значение свертки может быть записано как $\sum_{n=1}^k x_n y_{k-n+1}$, что в точности является k -й цифрой произведения xy .

Рисунок 7.1 является иллюстрацией к конкретному примеру. Первый столбец показывает реализацию умножения $25 \times 13 = 325$, выполняемого (по основанию 10) так, как этому обучают в начальной школе. Следует обратить особое внимание на то, что когда получается число, большее 9 (как в случае $3 \times 5 = 15$), то цифры старшего порядка переносятся и суммируются в соответствующих столбцах. Та же самая ситуация специально показана в среднем столбце для тех, кто не обращал ранее внимания на указанную процедуру. Операции переноса не были проделаны, а числа были накоплены в каждом из столбцов. Окончательный правильный ответ, однако, получается в том случае, если числа интерпретируют обычным образом, т. е. $2 \times 100 + 11 \times 10 + 15 \times 1 = 325$. Этот ответ представлен в записи со смешанным форматом. Последний столбец показывает свертку, выполняемую для последовательностей цифр у чисел 25 и 13. В данном примере число 13 перевернуто и сдвинуто (на дискретные шаги) за число 25. Перекрывающиеся столбцы перемножаются и суммируются. Величины, получающиеся на трех ступенях выполнения операции свертки, являются такими же, как и в среднем столбце, что подтверждает эквивалентность свертки и умножения. Процесс умножения с помощью операции свертки приобрел известность как алгоритм цифрового умножения с помощью аналоговой свертки (ЦУАС).

В целом для заданного основания b число представляют с точностью в l цифр, где каждая цифра изменяется от 0 до $b-1$. Если два числа, каждое длиной l цифр, умножаются

Умножение	Умножение без переноса	Свертка
25	25	25
<u>13</u>	<u>13</u>	<u>31</u>
75	615	15 → 15
<u>25</u>	25	25
325	<u>2115</u>	<u>31</u>
	•325	65 → 11
		25
		<u>31</u>
		2 → 2

Рис. 7.1. Три способа умножения (слева направо): обычная процедура, без переноса цифр в разрядах, свертка.

без арифметических переносов, то результат имеет длину $2l-1$ цифр, при этом каждая цифра лежит в интервале от 0 до $l(b-1)^2$. При электронном способе вычисления выражений предпочтительным является представление информации по основанию 2 (двоичный код). Это объясняется тем, что в данном случае высока помехозащищенность и можно применять булеву алгебру. Однако что касается оптических вычислений, то веских причин для выбора основания 2 нет. Это объясняется непрерывным характером прохождения света в различных средах. Наоборот, имеется соблазн рассмотреть другие, большие по величине основания с целью уменьшения объема данных. Как показано в [8], увеличение b позволяет снизить l ценой увеличения динамического диапазона выходного сигнала. Например, число, требующее 16 бит по основанию 2, потребует 8 бит по основанию 4 и 4 бит — по основанию 16. Динамический диапазон выходного сигнала составит 16 по основанию 2, 72 по основанию 4, и 900 по основанию 16. При переходе к основанию 16 был превышен достижимый на практике предельный динамический диапазон оптических систем (около 500). Однако даже в случае основания 4 начинает уменьшаться помехозащищенность. По этим причинам, а также по практическим соображениям обеспечения условий состыковки с электронными схемами дальнейшая дискуссия будет ограничена рассмотрением двоичных чисел и будет считаться, что в обозримом будущем оптические компьютеры будут оставаться двоичными.

7.2.2. Оптические устройства, выполняющие свертку при операциях умножения

Алгоритм ЦУАС оказался особенно привлекательным для исследователей, работающих в области оптики, вследствие той простоты, с которой операция свертки может быть выполнена оптически. Цель данной главы заключается в том, чтобы указать различные пути выполнения алгоритма ЦУАС, удовлетворяющего потребностям оптических вычислений. Эти потребности включают высокую пропускную способность, возможности расширения до задач более высокого уровня, приборную совместимость. Последнее требование представляет собой запутанный вопрос и будет сделана попытка лишь обрисовать его в целом. Очевидно, что если данная методика не может быть реализована в реальном времени в надежном, воспроизводимом и предпочтительно недорогом устройстве, то нет смысла рассматривать возможности ее применения. С другой стороны, если методика «на бумаге» выглядит многообещающей, это поможет направить исследования на развитие подходящей технологии для ее практической реализации. В данной главе в качестве исходной будет принята позиция, согласно которой любое выполняющее свертку устройство может быть реализо-

вано с помощью только лишь фотографического транспаранта, используемого в качестве устройства обработки данных в реальном времени, так что возможные архитектуры будут обсуждаться безотносительно к предельным возможностям конкретных приборов. Таким образом, удастся избежать морального устаревания информации и привязанности к одному конкретному устройству. В конце главы будут обсуждаться некоторые факторы, ограничивающие на практике возможности устройств и их влияние на развитие архитектур в будущем.

7.2.2.1. Операция свертки в области пространственных частот (фурье-свертка)

Имеются два способа выполнения операций свертки в оптике. Более общий метод состоит в выполнении свертки в фурье-плоскости. Преобразование Фурье одной функции осуществляется линзой, в то время как преобразование Фурье второй функции записано в голограмме. Для осуществления свертки выполняется обратное преобразование результата произведения фурье-образов. Используя эту методику, в [9] впервые была продемонстрирована работоспособность алгоритма двоичного ЦУАС.

Свертка, выполняемая в фурье-плоскости, демонстрирует многие из трудностей, присущих процессу нахождения удачного соотношения между конструкцией устройств и алгоритмами оптических вычислений. Сама по себе процедура выполнения ЦУАС для фурье-образов является крайне простой и быстрой. Стоит только расположить нужным образом входной пучок и голограмму, и операция свертки выполняется моментально. Существует ряд устройств, которые могли бы выполнять модуляцию входного сигнала с высокими скоростями (например, 10 МГц и более). Также существуют и фотодетекторы, обладающие таким же высоким быстродействием. Проблемы возникают с использованием голограмм. На момент написания данной книги не существует устройств, позволяющих производить электронную запись голограмм в реальном времени. Как будет пояснено ниже, из-за отсутствия таких голограмм теряется интерес к рассмотрению операций, выполняемых в фурье-плоскости, и больший интерес приобретают операции с временной и пространственной координатами. Но все же интересно чуть-чуть порассуждать о том, как происходила бы реализация операции свертки в плоскости фурье-преобразования, если бы существовало соответствующее устройство.

В данном случае рассмотрим случай умножения $6 \times 3 = 18$, или в двоичной записи $110 \times 011 = 100010$. Результат умножения в двоичной записи со смешанным форматом выглядит так: 01210. Входной сигнал в нашем случае может быть представ-

лен двумя дельта-функциями (одна для столбца четверок и одна для столбца двоек), разнесенных одна от другой на расстояние a . Фактически имеется третья дельта-функция, представляющая столбец единиц, но ее весовой коэффициент равен 0. На практике функция в каждой точке имела бы конечную ширину, но этот случай здесь не рассматривается. Обозначая пространственную частоту $\xi (=X/\lambda F$ для линзы с фокусным расстоянием F и длиной волны света λ), запишем соотношение между входным сигналом и плоскостью Фурье

$$\delta(X) + \delta(X - a) \xrightarrow{FT} 1 + e^{-i2\pi\xi a}.$$

Другими словами, фурье-плоскость состоит из набора плоских волн, углы распространения которых пропорциональны цифрам входного сигнала. При том же самом аргументе другое число представляется в фурье-плоскости выражением

$$\delta(X) + \delta(X + a) \xrightarrow{FT} 1 + e^{i2\pi\xi a}.$$

Перемножая два выражения и проводя преобразования, получаем

$$2 + e^{-2\pi\xi a} + e^{i2\pi\xi a} \xrightarrow{FT} 2\delta(X) + \delta(X - a) + \delta(X + a).$$

Для интерпретации данного результата следует принять во внимание, что столбец единиц расположен теперь в точке $-2a$; столбец двоек в точке $-a$ и т. д. до столбца, соответствующего разряду шестнадцать с координатой $+2a$. Видно, что при такой интерпретации результат составляет 01210, как и следовало ожидать.

Выполняя этот пример, можно дать иную интерпретацию методу выполнения свертки в фурье-плоскости и определить требования к приборному исполнению. Входной сигнал в плоскости Фурье, как упомянуто выше, состоит из плоских волн, направленных под разными углами. В фурье-плоскости имеется дифракционная решетка. В приведенном примере дифракционная решетка без изменений пропускает 50% света и отражает оставшиеся 50% на угол, описываемый величиной a . Другая линза собирает изменившие ранее свое направление световые лучи и направляет их на детектор. Теперь ясно, что не требуется взаимная когерентность источников входного сигнала; они должны быть только монохроматическими.

Среда для работающей в реальном времени голограммы должна обладать только способностью суммировать ограниченное число плоских волн. Можно представить, что акустооптическая ячейка могла бы обеспечить необходимую плотную дифракционную решетку. В данном случае, однако, не ставилась цель выполнить глубокую разработку конструкции, основываясь на приведенных соображениях, а скорее цель состояла в

том, чтобы проиллюстрировать необходимую степень взаимодействия между алгоритмами и технологией изготовления устройств. Алгоритмы должны реализовываться на доступных устройствах, но привлекательный алгоритм может явиться стимулом для развития подходящих устройств. Опыт, полученный на основе этих умозрительных экспериментов с конкретными алгоритмами и гипотетическим устройством, может быть использован для более детальной проработки реальных устройств.

7.2.2.2. Временная свертка

Альтернативой частотной свертке является операция свертки, выполняемая по временному аргументу. В обычной методике вычисления свертки преобразованная функция с обратной зависимостью от времени поддерживается постоянной, а относительно нее изменяется вторая функция. Их произведение во всех точках образует свертку. Временная свертка особенно привлекательна для оптических вычислений по двум причинам. Данная методика может выполняться на акустооптических брегговских ячейках, являющихся доступными, надежными и (относительно) недорогими. Может быть также использован ряд других подходов, включая интегральную оптику. Временная свертка также совместима с концепцией систолических процессоров [10]. В систолическом процессоре данные поступают в обрабатывающую ячейку, умножаются, суммируются и перемещаются в следующую обрабатывающую ячейку. Для получения общего результата вычислений складывают результаты из всех ячеек. Далее будет показано, что это именно тот способ, согласно которому работает устройство, выполняющее временную свертку.

На рис. 7.2 показаны два основных типа устройств, выполняющих временную свертку. В схеме с пространственным интегрированием (рис. 7.2, а) функция $g(x)$ вводится в устройство в виде пространственно изменяющейся функции (возможно, с помощью фиксированной маски или матрицы светодио-

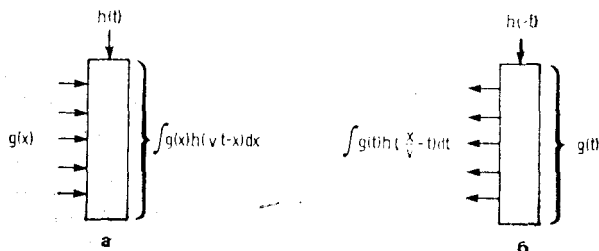


Рис. 7.2. Схемы выполнения свертки по области временной координаты: а — с пространственным интегрированием; б — с временным интегрированием,

дов), а само устройство работает как сдвиговый регистр. В качестве такого устройства могут применяться акустооптическая ячейка, устройство на поверхностных акустических волнах или даже ПЗС-сдвиговый регистр, выполняющий дискретный сдвиг функции. Функция $g(x)$ поддерживается постоянной, в то время как $h(t)$ сдвигается относительно нее. Свет, выходящий из сдвигового регистра, суммируется по пространству с помощью линз (изображенных скобками) на одном приемнике. Тогда временная зависимость $f(t)$ сигнала, выходящего из детектора, представляет собой функцию

$$f(t) = \int g(x) h(vt - x) dx,$$

являющуюся сверткой g и h . На рис. 7.2, б показана схема временного интегрирования. Одна из функций, $g(t)$, подается в качестве зависящего от времени сигнала на источник света, например светодиод или полупроводниковый лазер. Другая функция, $h(-t)$, подается на сдвиговый регистр. Функция $g(t)$ равномерно «распределяется» по сдвиговому регистру с помощью линзы (снова изображенной скобками), в то время как $h(t)$ сдвигается относительно нее. Свет, выходящий из сдвигового регистра, суммируется по времени дискретными элементами или матрицей фотодетекторов. Сигнал в матрице фотодетекторов можно представить в виде

$$f(x) = \int g(t) h\left(\frac{x}{v} - t\right) dt,$$

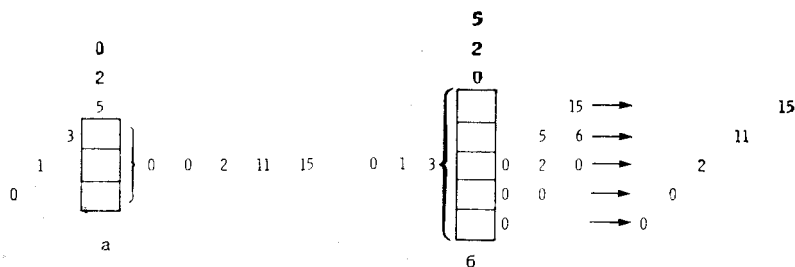
где v — скорость сдвига функции h в сдвиговом регистре. Сигнал $f(x)$ представляет вариант пространственной выборки операций свертки от g и h .

При рассмотрении рис. 7.2 становится очевидным, что устройства, производящие пространственную и временную свертки, являются зеркальными образцами друг друга, где выходной и входной сигналы поменялись ролями. Для умножения двух чисел может быть использована любая из конфигураций, при этом функции g и h представляют в виде последовательностей цифр, описывающих соответствующие числа. Свертка должна выполняться дискретно; данные фотодетекторов могут считываться только тогда, когда цифры, представляющие g , синхронизированы с цифрами, описывающими h . Это требует импульсного режима работы источника света или детектора. Для наглядности представим сдвиговые регистры как строки из клеточек, по которым перемещаются цифры, пробегающие по одной клеточке за тактовый цикл. Все входные и выходные функции будут записаны аналогичным образом, чтобы показать, каким образом данные должны быть форматированы. Ни-

же снова будет рассмотрен пример умножения десятичных чисел $13 \times 25 = 325$.

На рис. 7.3 показана схема вычисления свертки с пространственным и временным интегрированием для выполнения операции умножения. В схеме с пространственным интегрированием (рис. 7.3, а) сдвиговый регистр состоит из трех ячеек (так как в данном примере $l=3$). Число 013 передвигается по регистру синхронно с цифрами числа 025 по регистру. Если цифра вводится в систему источником света, то она сохраняется в ней в течение трех циклов, в то время пока другое число перемещают относительно него. Для числа 025 требуются пять циклов, чтобы полностью очистить сдвиговый регистр и освободить место для умножения следующего числа. Таким образом, за этим числом должны последовать два буферных нуля, которые также оказываются включенными в последовательность цифр выходного сигнала (следует обратить внимание на смешанный формат выходного сигнала). Эти два буферных «пространства» рассчитаны на случай переполнения, которое может возникнуть после преобразования результата в обычный, несмешанный формат представления числа. В течение соответствующего буферного интервала времени источники входного светового сигнала могут быть выключены на период, составляющий два тактовых цикла. Этот интервал времени может быть использован для медленного переключения входного сигнала от светодиодов из одного значения в другое. В этом случае источник света может иметь гораздо большее время срабатывания, чем тактовый цикл. Фотодетекторы, однако, должны реагировать на каждый тактовый цикл.

В случае временного интегрирования (рис. 7.3, б) для вычисления полной свертки требуется наличие пяти ячеек в сдвиговом регистре. Теперь для очистки регистра от числа 025 требуется семь тактовых циклов. Следует заметить, что цифры 025 вводятся в обратном порядке по сравнению со случа-



Примечание: $2 \quad 11 \quad 15 = 2 \times 10^2 + 11 \times 10^1 + 15 \times 10^0 = 325$

Рис. 7.3. Схема умножения $13 \times 25 = 325$, выполняемого с помощью операции дискретной свертки цифр: а — с пространственным интегрированием; б — с временным интегрированием.

ем пространственного интегрирования. Число должно быть полностью загружено в регистр до начала выполнения операции свертки, так что числу 013 должны предшествовать два буферных нуля (показанные в выходном сигнале). За числом 025 также должны следовать два буферных нуля (также показанные в выходном сигнале). Входной источник света должен отслеживать каждый тактовый импульс, что требует от источника высокого быстродействия. Фотодетектор интегрирует сигнал на протяжении трех циклов. Результат интегрирования периодически выводится на матрицу фотодетекторов. Это позволяет также периодически производить считывание данных с детекторов за то время, пока выполняется умножение. В другом варианте архитектуры с временным интегрированием сдвиг обоих чисел относительно друг друга производится в противоположных направлениях, что дает сокращение времени, вдвое затрачиваемого на выполнение свертки. Такая архитектура является совместимой с методами интегральной оптики [8], однако она не пригодна для применения в систолических системах и далее не будет обсуждаться.

Два основных блока умножителей, показанные на рис. 7.3, могут использоваться в различных сочетаниях при формировании систолических процессоров. Процессоры данного типа ранее были описаны в [11], и ниже это описание просто изложено повторно. Блоки умножителей могут соединяться как последовательно, так и параллельно. В первом случае имеется один входной сигнал h и несколько входных сигналов g . Соответственно в случае умножения матрицы на вектор элементы вектора представляют собой входной сигнал h , а элементы каждой строки матрицы являются входными сигналами g . Эта схема требует сдвига только в одном измерении и будет далее именоваться одномерной архитектурой.

Если блоки умножителя размещаются параллельно, то тогда имеется лишь один входной сигнал g , который необходимо перемещать относительно нескольких входов h . Соответственно g представляет входной вектор, а h — строки матрицы. Поскольку требуется сдвиг в двух измерениях, то далее такая архитектура называется двумерной.

Ниже представлены примеры выполнения свертки каждым из типов архитектур, рассмотренных для одной и той же задачи. Для большей убедительности взят случай умножения матрицы на вектор. Задача выглядит следующим образом.

$$\begin{pmatrix} 001 & 100 \\ 011 & 110 \end{pmatrix} \begin{pmatrix} 101 \\ 010 \end{pmatrix} = \begin{pmatrix} 00101 + 01000 \\ 01111 + 01100 \end{pmatrix} = \begin{pmatrix} 01101 \\ 11011 \end{pmatrix}$$

A
 b
 c

Хотя в приведенном примере не используется смешанный формат, но в результате выполнения задачи возникает пере-

полнение. Кроме того, в приведенной задаче используются лишь целые числа. Однако, как показал опыт, полученный при эксплуатации ЭВМ, двоичное представление с успехом может быть использовано для обозначения различных знаков, дробей и даже комплексных чисел (хотя и за счет необходимости использования дополнительных разрядов). В процессе проверки того, как каждый процессор выполняет указанную задачу, имеется возможность сделать общие выводы о требованиях, предъявляемых к интерфейсным устройствам и скорости обработки.

Одномерная архитектура. Сперва рассмотрим одномерную архитектуру с пространственным интегрированием (рис. 7.4). Ранее эта архитектура была описана в [12] и получила известность как схема оптического систолического матричного процессора (ОСМП). Для умножения матрицы $m \times n$ на вектор $n \times 1$ с точностью в l цифр ОСМП должен иметь один вход для ввода вектора, ml входов для ввода матричных элементов и m детекторов выходного сигнала. Элементы c , представляющие собой изменяющиеся по времени сигналы от m детекторов, подаются параллельно. Для решения задачи эти сигналы требуется просуммировать и перезаписать в смешанном формате. Последующий сигнал показывает, что ОСМП использует $(2n+m-1)l-n$ тактовых цикла для выполнения операции умножения.

На рис. 7.5 показана одномерная архитектура с временным интегрированием. Как для элементов A , так и для элементов b требуются буферные разряды. В схеме имеется m входов для матричных элементов и один вход для ввода вектора. Для

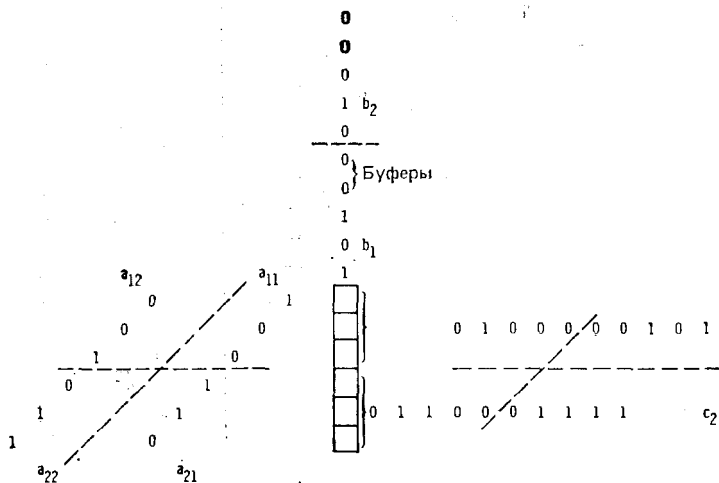


Рис. 7.4. Схема одномерного умножителя матрицы на вектор с пространственным интегрированием.

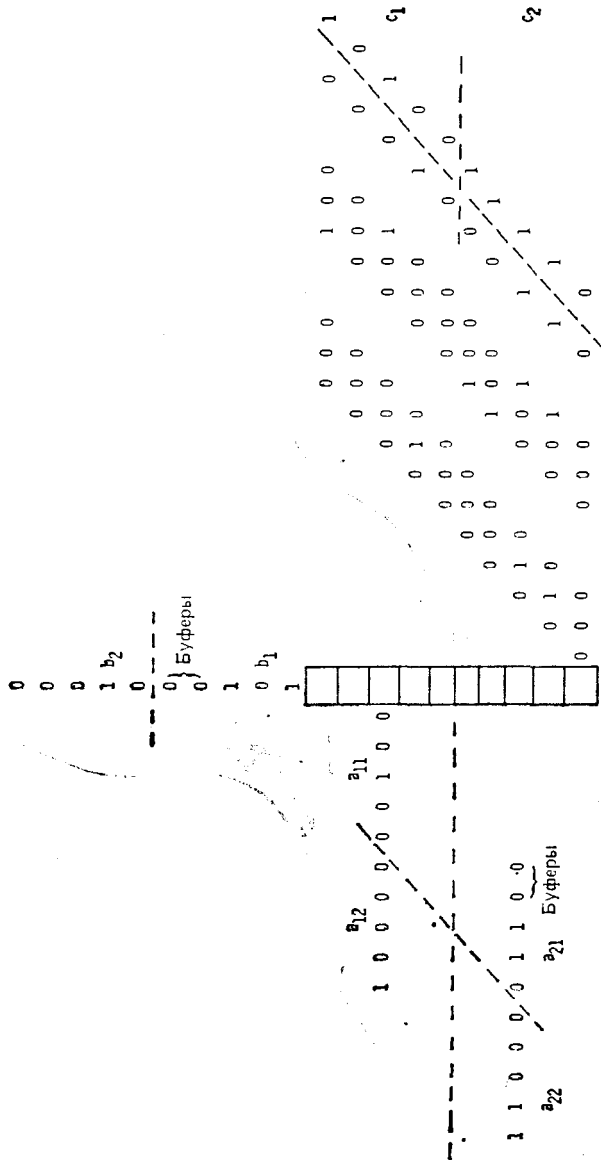


Рис. 7.5. Схема одномерного умножителя матрицы на вектор с временным интегрированием.

регистрации выходного сигнала требуется $(2l-1)m$ фотодетекторов. Выходной сигнал может быть представлен в виде последовательности определенных сигналов путем соответствующего переключения детекторов, относящихся к заданным элементам вектора выходного вектора. Получающаяся последовательная серия сигналов может передаваться в соответствующую схему для выполнения суммирования и преобразования результата из смешанного формата в обычный. Процессор указанного вида требует использования m таких же параллельных схем, как в случае ОСМП. Для выполнения процедуры умножения матрицы на вектор требуется $(m+n-1)(2l-1)$ тактовых циклов.

В [13] описан вариант одномерной архитектуры с временным интегрированием, позволяющий резко увеличить скорость вычисления. Это достигается путем увеличения числа входов до величины ml и сдвига буферных нулей между входными сигналами. Эта процедура также предусматривает сдвиг цифр, возникающих при переполнении, от выходного сигнала (рис. 7.6). Теперь для завершения процедуры матрично-векторного умножения требуется лишь $(m+n-1)l$ тактовых циклов.

Двумерная архитектура. В [14] описана двумерная архитектура с пространственным интегрированием и рассмотрен случай систолического акустооптического двоичного процессора выполнения свертки (САОДПС). Как показано на рис. 7.7, для САОДПС требуется l входов для ввода вектора и m входов для ввода матрицы. Имеется m детекторов выходного сигнала. САОДПС должен иметь два набора сдвиговых регистров; один набор быстро сдвигает матричные элементы относительно элементов вектора, в то время как другой набор регистров медленно передвигает вектор последовательно по всем строкам матрицы. Вектор выходного сигнала представляет собой m последовательных серий, которые требуется просуммировать и преобразовать. Процесс умножения занимает время, равное $(n+m-1)(2l-1)$ тактовых циклов. Последний из обсуждаемых умножителей матриц на вектор, показанный на рис. 7.8, представляет двумерную архитектуру с временным интегрированием. Имеется один вход для вектора, который требуется развернуть и переместить относительно m входов матрицы. Выходной сигнал требует $(2l-1)m$ фотодетекторов, которые должны быть синхронизированы с целью параллельного вывода выходных сигналов. Затраты времени составят в этом случае $n(2l-1) + (m-1)$ тактовых циклов.

Умножитель матрицы на матрицу. На основе описанного выше умножителя на вектор может быть создан умножитель матрицы на матрицу. Например, k одномерных умножителей с пространственным интегрированием могут быть размещены параллельно для умножения матрицы $m \times n$ на матрицу $n \times k$.

Аналогичная конфигурация возникает при последовательном размещении двумерных умножителей с пространственным интегрированием. Очевидно, что при таком методе конструирования могут быть получены только два вида умножителей матрицы на матрицу, а именно варианты с пространственным и временным интегрированием. Оба они показаны ниже, но для простоты изображены лишь входные сигналы. Выходные сиг-

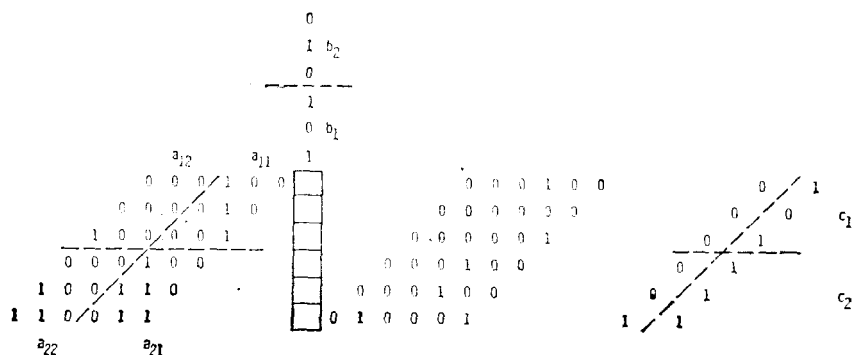


Рис. 7.6. Схема одномерного умножителя матрицы на вектор с временным интегрированием, обладающего повышенным быстродействием [13].

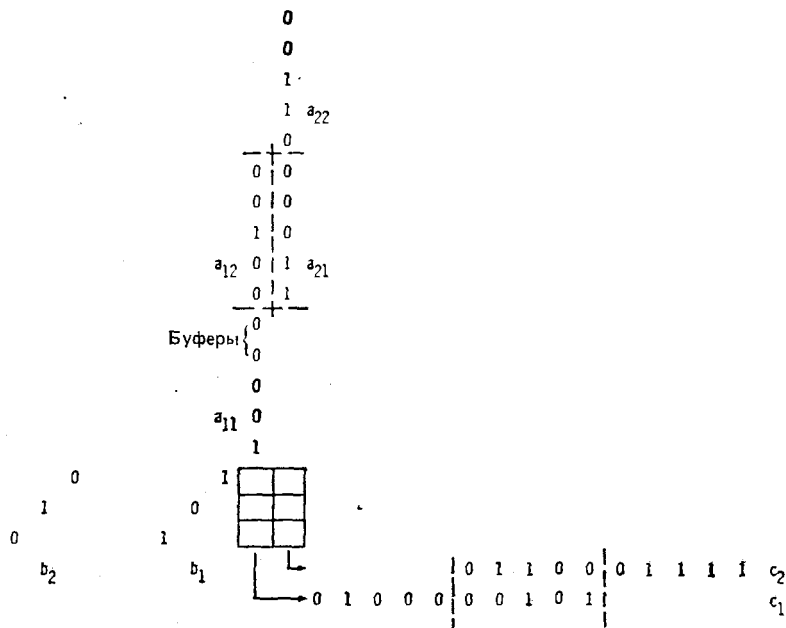


Рис. 7.7. Схема двумерного умножителя матрицы на вектор с пространственным интегрированием [14].

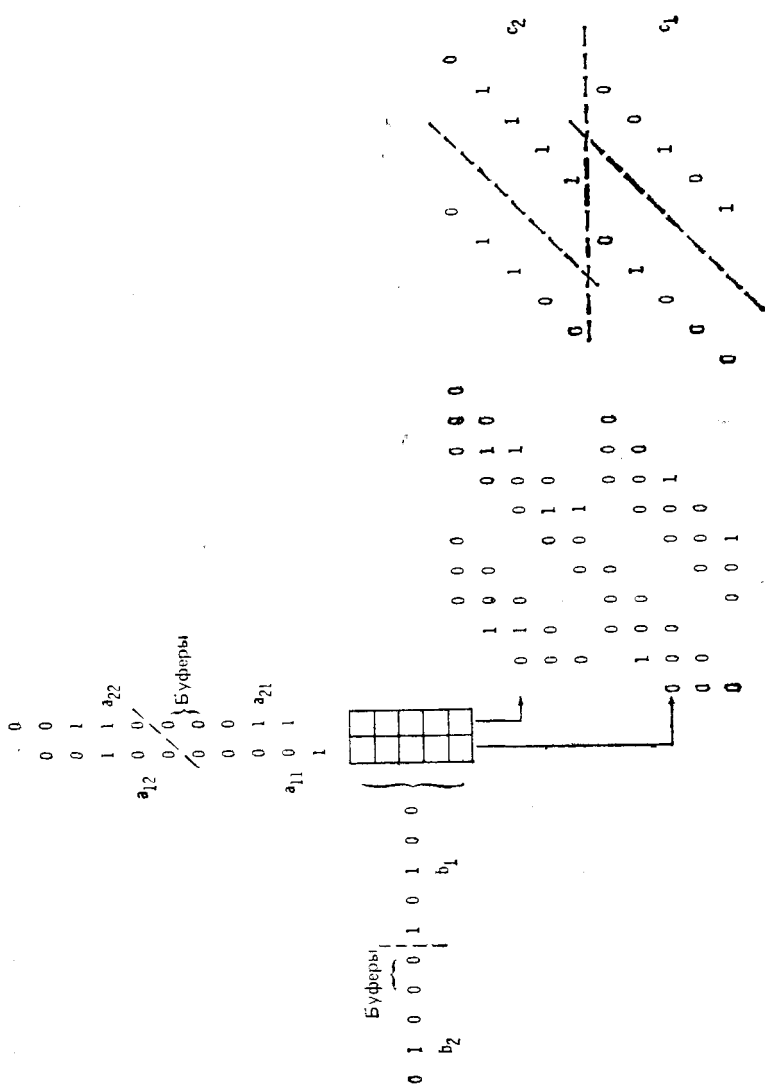


Рис. 7.8. Схема двумерного умножителя матрицы на вектор с временным интегрированием.

налы могут быть получены с помощью соответствующих схем умножителей матрицы на вектор.

На рис. 7.9 показан вариант с пространственным интегрированием, где матричное произведение есть AB при $b_{11}=101$, $b_{12}=010$, $b_{21}=011$, $b_{22}=001$. Имеется k входов для одной матрицы и ml входов для другой. Для выходной матрицы имеется mk фотодетекторов. Вид входных сигналов позволяет заключить, что выполнение всей процедуры потребует $(2m+n-1)l-n+(k-1)(2l-1)$ тактовых циклов. Показанный на рис. 7.10 вариант схемы с временным интегрированием требует наличия k входов для одной матрицы и m входов для другой. Имеется $(2l-1)mk$ синхронизируемых выходов. Операция умножения занимает лишь $(m+n-1)(2l+1)+(k-1)$ тактовых циклов. Соответственно архитектура с временным интегрированием допускает большую степень параллелизма при умножении матрицы на матрицу.

Более совершенные характеристики были достигнуты для устройства, названного его создателем «быстродействующим биполярным некогерентным вычислительным устройством на светоделительном кубе, работающим без подачи напряжения», или сокращенно RUBIC cube [15] (рис. 7.11). Как и ранее, улучшение было достигнуто за счет увеличения числа вхо-

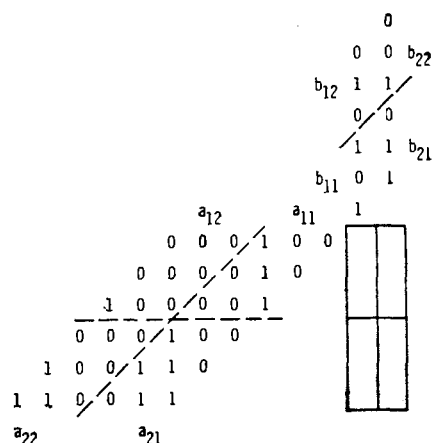
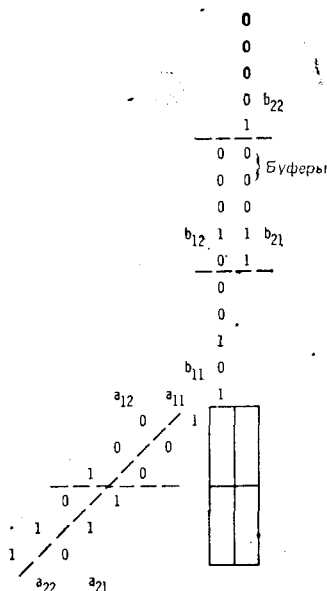


Рис. 7.9. Схема умножителя матрицы на матрицу с пространственным интегрированием.

Рис. 7.10. Схема умножителя матрицы на матрицу с временным интегрированием.

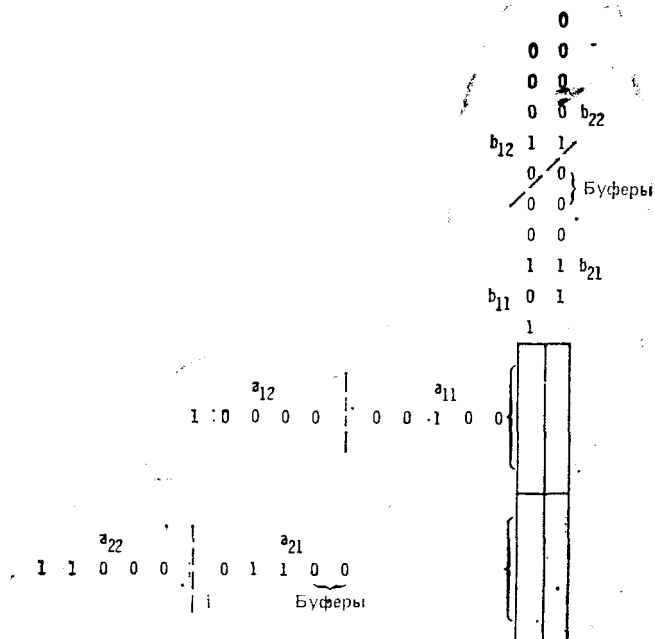


Рис. 7.11. Схема умножителя матрицы на матрицу с временным интегрированием, обладающего повышенным быстродействием [15].

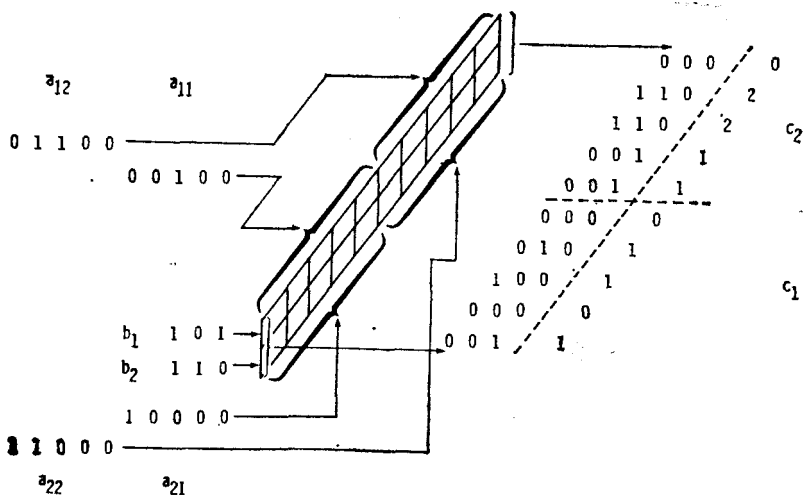


Рис. 7.12. Схема умножителя матрицы на вектор с временным интегрированием [16].

дов и исключения возможностей возникновения с переполнением регистров. Для одной матрицы требуется k входов, для другой ml входов. Имеется lmk выходов. Полное время операции вычисления составляет лишь $(n+m-1)l + (k-1)$ тактовых циклов.

Разновидности основной архитектуры. Сообщалось и о других способах преобразования схем вычисления свертки в схемы умножителей матрицы на матрицу. В [16] для получения промежуточного произведения при вычислении внутреннего произведения двух векторов используется основная схема вычисления свертки с интегрированием по времени. Все промежуточные произведения вычисляются параллельно на независимых друг от друга умножителях и суммируются с помощью цилиндрической линзы. Таким образом, для перемножения двух векторов, состоящих из n элементов, с точностью в l знаков требуется n входов для каждого вектора, $2l-1$ фотодетекторных элементов и $2l-1$ тактовых циклов. При выполнении суммирования с помощью линз максимальное значение на детектирующем элементе составляет $nl(b-1)^2$. Матрично-векторный умножитель схематично показан на рис. 7.12. Следует заметить, что буферные нули в данном случае не требуются, поскольку элементы вводятся параллельно. Для построения матрично-векторного умножителя для перемножения матрицы $m \times n$ и вектора $n \times 1$ все m умножителей векторов размещаются параллельно. Теперь каждый элемент матрицы a имеет вход (при общем числе входов mn), а элементы вектора b сдвигаются относительно этих входов. Умножение выполняется за интервал времени, составляющий $m(2l-1)$ циклов; при этом используется $m(2l-1)$ детекторов выходного сигнала. Возможности процессора удается расширить до операции умножения матрицы на матрицу с помощью временного разделения каналов для ввода элементов b при условии строчной загрузки матрицы по соответствующим буферам. В схеме имеется также mn входов для одной матрицы и n входов для другой, а также $m(2l-1)$ детекторов выходного сигнала. Затраты времени на вычисления составляют $(k+m-1)(2l-1)$ тактовых циклов.

В работе [17] описан матрично-векторный умножитель с улучшенными возможностями, в котором используется частотное разделение каналов в акустооптической ячейке. Как указано в [17], данная архитектура не может быть незамедлительно приспособлена для цифровых вычислений, но ее можно несколько модифицировать для обработки цифровых данных. Вначале рассмотрим схему умножителя вектора на вектор, выполняющего свертку с пространственным интегрированием. Для достижения точности вычислений в l цифр при размерности вектора n необходимо использовать l входов для ввода одного вектора и один вход для ввода другого вектора, а обрабатывающая

ячейка должна иметь l разрядов. В схеме имеется один фотодетектор выходного сигнала; для вычисления внутреннего произведения требуется $n(2l-1)$ тактовых циклов. Если в качестве входного сигнала в акустооптической ячейке рассматривается одна строка матрицы A , то все строки A могут быть введены одновременно путем использования различных несущих частот для каждой из строк. Следовательно, все элементы матрицы выходного сигнала появляются в выходной плоскости одновременно (рис. 7.13). Различные элементы c несколько смещены на рисунке, поскольку линза преобразует частоту входного сигнала в координату выходного сигнала. В данном случае элементы вектора b все еще требуют наличия l входов, в схеме также

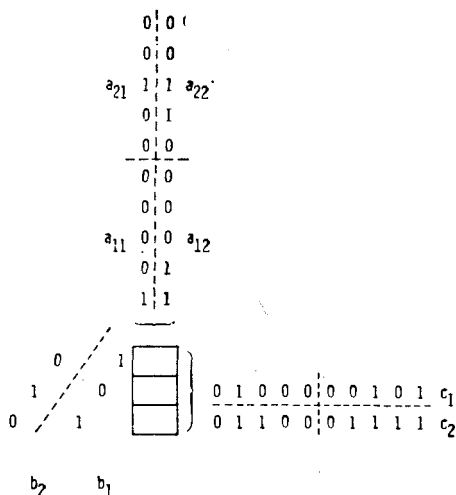


Рис. 7.13. Умножитель матрицы на вектор с частотным разделением каналов [17].

имеется m детекторов выходного сигнала. Процедура матрично-векторного умножения производится за $n(2l-1)$ тактовых циклов. Длительность тактового цикла не зависит от m , но величина m ограничена возможностями акустооптической ячейки (см. разд. 7.3.2).

7.2.2.3. Способы определения внешнего произведения методом вычисления свертки

Все обсуждавшиеся до сих пор методы позволяют умножать две матрицы или матрицу на вектор путем выполнения операции внутреннего произведения между строками одной матрицы и столбцами другой. Согласно [18], произведение матриц может также быть найдено с помощью вычисления внешнего про-

изведения. Внешнее произведение определяется выражением $a \times b^T = c$, где a — m -мерный вектор, b — n -мерный вектор, а c — матрица $m \times n$. Для частного случая с $m = n = 3$ получаем

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times (b_1 \ b_2 \ b_3) = \begin{pmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 \\ a_2 b_1 & a_2 b_2 & a_2 b_3 \\ a_3 b_1 & a_3 b_2 & a_3 b_3 \end{pmatrix} = c.$$

Если элементы c суммируются вдоль противоположных диагоналей, то может быть получен новый вектор c' , компоненты которого

$$c' = (a_1 b_1, a_2 b_1 + a_1 b_2, a_3 b_1 + a_2 b_2 + a_1 b_3, a_3 b_2 + a_2 b_3, a_3 b_3)$$

представляют собой не что иное, как свертку a и b . Отсюда следует, что если a и b являются цифровыми представлениями чисел, то можно определить результат их перемножения.

Матрично-векторное произведение выполняется путем вычисления внешнего произведения первого столбца матрицы и первого элемента вектора, сложения с результатом внешнего произведения второго столбца и второго элемента и т. д.

Например, произведение

$$\begin{pmatrix} 1 & 4 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 13 \\ 27 \end{pmatrix}$$

в двоичной записи выглядит так;

$$\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \times (1 \ 0 \ 1) + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \times (0 \ 1 \ 0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Если две подматрицы соответствующей матрицы суммируются вдоль противодиагоналей, то в результате получим

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 0 & 2 & 2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 13 \\ 27 \end{pmatrix}.$$

Возможности умножителя могут быть легко расширены для случая умножения матрицы на матрицу.

Для больших матриц, разделенных на подматрицы, процедура суммирования вдоль противодиагоналей может оказаться весьма тяжелой задачей. В работе [19] было предложено выполнять суммирование вдоль противодиагоналей оригинальным

способом, записывая вектор в виде матрицы следующим образом:

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \times$$

$$\times \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 2 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Правильный результат получают уже в смешанном формате, путем умножения исходной матрицы на дополненный вектор. Эта процедура может быть выполнена с помощью любого из обсуждавшихся выше вариантов умножения матрицы на матрицу. Ценой этого будет служить увеличение объема памяти, необходимое для записи вектора, и задача класса умножение матрицы на вектор оказывается «отданной в уплату» за задачу класса матрица — матрица. В целом же кажется более выгодным использовать оптические методы для выполнения суммирования вдоль противодиагоналей. В число возможных оптических способов решения такой задачи входит сегментация цилиндрических линз, сегментация голографических линз или применение матриц оптических волокон, соединяющих соответствующие элементы с детекторами, на которых происходит суммирование.

В работе [20] предложены две возможные схемы построения процессоров внешнего произведения. В первом случае используется перекрестное включение одномерных входных модуляторов (рис. 7.14). Для умножения матрицы на вектор в один из модуляторов вводят целый столбец матрицы, а элементы вектора размещаются в другом модуляторе. Матричный модулятор должен обладать ml разрядами, а модулятор для ввода вектора должен иметь l разрядов. Когда оба модулятора загружены, то от источника света подается импульс света и перекрестное произведение записывается на матрице $ml \times l$ интегрирующих по времени детекторов. Если суммирование осуществляется оптически, необходимо только $m(2l-1)$ детекторов. Каждое промежуточное произведение может быть «накоплено» на детекторе за время загрузки входного сигнала в модулятор, которое полагаем равным ml . Полное число тактовых импульсов для операции умножения матрицы на вектор составляет nml . Для умножения матрицы на матрицу требуется $km(2l-1)$ детекторов, при этом необходимое число тактовых циклов составляет лишь nml (если $m \geq k$).

Во втором методе используется один двумерный пространственный модулятор света, адресация которого проводится с по-

Таблица 7.1. Рабочие характеристики умножителей матрицы на вектор

Тип	Число тактовых циклов на операцию умножения	Время ¹⁾ , мкс ($l=16, n=m=32$)	Мега·оп/с ²⁾	Время, мс ($l=32, n=m=128$)	Мега·оп/с
1М ПИ	$(2n+m-1)l-n$	148,8	13,8	1,213	31,1
1М ВИ (I)	$(m+n-1)(2l-1)$	195,3	10,5	1,607	23,5
1М ВИ (II)	$(m+n-1)l$	100,8	20,3	0,816	46,2
2М ПИ	$(m+n-1)(2l-1)$	195,3	10,5	1,607	23,5
2М ВИ (I)	$n(2l-1) + (m-1)$	102,3	20,0	0,819	46,0
2М ВИ (II)	$m(2l-1)$	99,2	20,6	0,806	46,7
ЧМ	$n(2l-1)$	99,2	20,6	0,806	46,7
1М ВВП	ml	51,2	40,0	0,410	92,0
2М ВВП	$2n$	6,4	320,0	0,026	1472

¹⁾ Тактовая частота считается равной 10 МГц.

²⁾ оп/с — операций в секунду; число операций = $2ml$.

мощью специальной перекрестной матрицы согласно данным, показанным на рис. 7.15. Если предположить, что возможна параллельная адресация всех перекрещивающихся электродов, тогда адресация всей матрицы может быть осуществлена за один тактовый цикл. В этом случае за время другого тактового цикла можно включить источник света и произвести запись набора величин промежуточных произведений. Для умножения матрицы на вектор необходимо $ml \times l$ модуляторов, $m(2l-1)$ фотодетекторов, процедура занимает $2n$ тактовых циклов. Для умножения матрицы на матрицу требуется $ml \times kl$ модуляторов, $kn(2l-1)$ фотодетекторов, затраты времени составляют лишь $2n$ тактовых циклов.

7.3. Рабочие характеристики

В предыдущих разделах описан ряд схем умножителей матрицы на вектор и матрицы на матрицу. Ниже представлен список этих умножителей, где каждая из схем получила специальное обозначение, используемое ниже в таблицах с 7.1 по 7.4.

Умножители матрицы на вектор

1М ПИ — одномерный с пространственным интегрированием

1М ВИ (I) — одномерный с временным интегрированием

1М ВИ (II) — одномерный с временным интегрированием [13]

2М ПИ — двумерный с пространственным интегрированием

2М ВИ (I) — двумерный с временным интегрированием

2М ВИ (II) — двумерный с временным интегрированием [16]

1М ВП — вычисляющий внешнее произведение с помощью одномерных модуляторов

2М ВП — вычисляющий внешнее произведение с помощью двумерных модуляторов

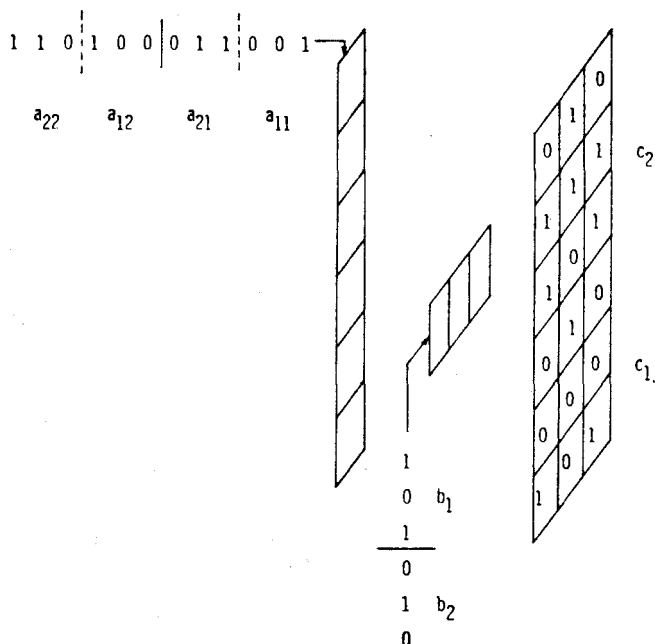


Рис. 7.14. Умножитель матрицы на вектор, вычисляющий внешнее произведение для одномерного входного сигнала. Накопленные результаты суммируются вдоль противодиагоналей подматриц размером 3×3 [20].

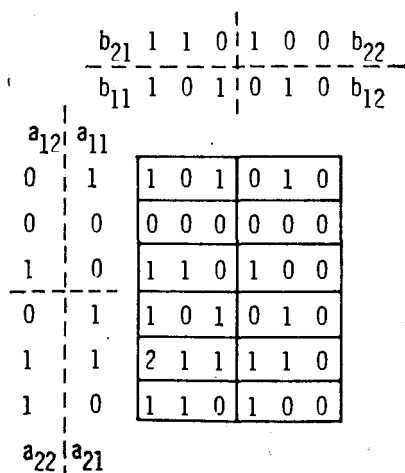


Рис. 7.15. Умножитель матрицы на матрицу, вычисляющий внешнее произведение для двумерного входного сигнала. Накопленные результаты суммируются вдоль противодиагоналей подматриц размером 3×3 [20].

Таблица 7.2. Рабочие характеристики умножителей матрицы на матрицу

Тип	Число тактовых циклов на операцию умножения	Время ¹⁾ , мс ($l=16, n=m=32$)	Мега·оп/с ²⁾	Время, мс ($l=32, n=m=k=128$)	Гига·оп/с
ПИ	$(2n+m-1)l - n + (k-1)(2l-1)$	0,2449	268	2,013	2,09
ВИ (I)	$(m+n-1)(2l-1) + (k-1)$	0,1984	330	1,519	2,59
ВИ (II)	$(n+m-1)l + (k-1)$	0,1039	631	0,829	5,06
ВИ (III)	$(k+m-1)(2l-1)$	0,1953	336	1,607	2,61
1М ВМП	nml	1,6384	40	52,429	0,08
2М ВМП	$2n$	0,0064	10240	0,026	164

¹⁾ Тактовая частота считается равной 10 МГц.

²⁾ оп/с — число операций в секунду: число операций = $2mnk$.

Таблица 7.3. Отношение Псалтиса для умножителей матрицы на вектор

Тип	$l=16, n=m=32$	$l=32, n=m=128$
1М ПИ	0,022	0,011
1М ВИ (I)	0,008	0,004
1М ВИ (II)	0,016	0,008
2М ПИ	0,016	0,008
2М ВИ (I)	0,016	0,008
2М ВИ (II)	0,017	0,008
ЧМ	0,032	0,016
1М ВМП	0,032	0,016
1М ВМП	0,258	0,250

Таблица 7.4. Отношение Псалтиса для умножителей матрицы на матрицу

Тип	$l=16, n=m=k=32$	$l=32, n=m=k=128$
ПИ	0,013	0,006
ВИ (I)	0,008	0,004
ВИ (II)	0,031	0,015
ВИ (III)	0,271	0,518
1М ВМП	0,001	0,0001
2М ВМП	0,258	0,250

Умножитель матрицы на матрицу

ПИ — с пространственным интегрированием

ВИ (I) — с временным интегрированием

ВИ (II) — с временным интегрированием [15]

ВИ (III) — с временным интегрированием [16]

1М ВП — вычисляющий внешнее произведение с помощью одномерных модуляторов

2М ВМ — вычисляющий внешнее произведение с помощью двумерных модуляторов

Цель данного раздела состоит в сравнении всех этих архитектур с определенных общих позиций. Чтобы сделать это, требуется дать определение целому ряду параметров, однако всегда имеется опасность сделать это слишком формальным, отдаленным от практики способом. С другой стороны, следует избегать опасности заняться научными «спекуляциями», аккуратно подгоняя сами параметры с целью демонстрации идеальных характеристик для любой из архитектур. Спорным является также и вопрос о том, может ли какая-либо из этих архитектур при определенных условиях работать лучше, чем указано в представленных ниже таблицах. Во всяком случае условия, при которых осуществляется такое улучшение, должны быть описаны очень тщательно.

Представленные здесь таблицы наиболее полезны для сравнительных оценок архитектур. Так как сравнение производится в общем виде, ни одной из архитектур не может быть приписано преимущество над другими. Тем не менее читателям, интересующимся возможностью использования тех или иных архитектур для конкретных задач, можно рекомендовать ознакомиться здесь с основными характеристиками и в дальнейшем расширить свои знания уже применительно к своим собственным потребностям.

7.3.1. Быстродействие процессора

В табл. 7.1 приведена скорость выполнения операции умножения матрицы на вектор. В первой колонке представлены выражения для числа тактовых циклов, необходимых для завершения одной операции умножения. Умножение матрицы $m \times n$ на вектор $n \times 1$ требует $2mn$ операций сложения и умножения. Если предположить, что биты данных проходят в системе с частотой 10 МГц (величина 0,1 мкс/бит является достаточно обоснованной для существующих электронных устройств), то можно вычислить скорость выполнения операций. Представлены два случая. Первый из них соответствует $n = m = 32$, при $l = 16$ (l эквивалентно точности вычислений), а второй случай относится к $n = m = 128$, $l = 32$. В табл. 7.2 представлены аналогичные данные для умножителей, выполняющих умножение матрицы на матрицу с точностью l цифр. В третьем столбце показаны результаты для $l = 16$, $n = m = k = 32$, а четвертый столбец соответствует $l = 32$, $n = m = k = 128$. Во всех случаях результаты даны для операций с фиксированной запятой, выполняемых в одну секунду. Ни один цифровой процессор (оптический или элект-

ронный) не работает естественным образом с плавающей запятой, так же как с различающимися по знаку или комплексными числами. Все эти возможные случаи должны выполняться с помощью специального кодирования (уменьшающего достигаемую на практике точность вычислений для заданного l) и параллельной обработки положительных и отрицательных действительных и мнимых чисел.

Все процессоры характеризуются тем, что чем больше объем задачи, тем быстрее они работают. Например, умножители матрицы на вектор (за одним исключением) работают со скоростью в десятки мегаопераций в секунду, характерной для небольших по объему задач. Процессоры умножения матрицы на матрицу работают со скоростями порядка гигаопераций в секунду, что характерно для задач большего масштаба. Очевиден выигрыш в быстродействии, получаемый для параллельной обработки. Более неопределенной характеристикой является абсолютная величина быстродействия. Представленные здесь цифры являются лишь оценками, но они действительно отражают общие свойства оптических вычислений. За небольшим числом исключений имеется очень небольшая разница (менее чем на порядок по величине) между разными видами оптических процессоров, используемых для решения заданной задачи. Возможности оптических процессоров, как представляется, достаточно жестко ограничены определенными скоростями. При сравнении с возможностями электронных процессоров скорости вряд ли произведут на читателя глубокое впечатление. На момент написания книги на промышленно освоенных электронных устройствах удается достичь скоростей около 50 мегаопераций в секунду. Разрабатываемые в настоящее время умножители матрицы на вектор не позволяют превзойти эту величину. Умножители матрицы на матрицу демонстрируют существенно более высокие возможности.

7.3.2. Проблемы разработки систем и возможности их приборной реализации

На пути создания успешно работающего оптического умножителя матриц возникает большое число различных проблем. Задача может быть условно разделена на проблемы, связанные с оптической частью, и проблемы, связанные с электронной частью. Проблемы оптических устройств, рассматриваемые здесь, затрагивают в основном вопросы подбора источников света для выходного сигнала, пространственных модуляторов света и фотодетекторов выходного сигнала. Проблемы, связанные с электронными устройствами, включают передачу данных на световые источники и модуляторы, а также восстановление данных, поступающих с детекторов. Эти проблемы настолько

обширны, что не могут быть здесь рассмотрены. Приборные решения также зависят от доступной технологии, которая в свою очередь сама быстро развивается. Следует только упомянуть две проблемы, наиболее тесно связанные с оптикой. Первая — это скорость передачи данных. Для определенности будем полагать, что электронные схемы передают информацию со скоростями около 10×10^6 бит/с (что будет соответствовать мегагерцевым частотам). Такая скорость обработки совместима с интегральными схемами со сверхвысокой степенью интеграции (СБИС). Любой источник света, фотодетектор или модулятор должны устойчиво работать с такой скоростью. При обсуждении полосы пропускания будет использоваться эмпирическое соотношение, согласно которому ширина полосы частот должна быть в четыре раза больше скорости передачи данных. Теорема Уиттекера — Шеннона¹⁾ требует различия этих параметров минимально в два раза, так что коэффициент 4 создает «резерв» для большей надежности на случай пиковых изменений уровня сигнала. Коэффициент 10 определяется результатами более жестких оценок, полученных из общих соображений, но далее будет показано, что и меньший множитель является приемлемым. Также было сделано предположение, что электронные схемы могут передавать данные по необходимости либо в последовательном, либо в параллельном виде. На практике возможная степень параллельности обработки может быть ограничена. Как показано ниже, это жестко ограничивает возможности двумерных процессоров, вычисляющих внешнее произведение.

Другая проблема, являющаяся гораздо более жесткой, заключается в необходимости для аналого-цифрового преобразователя (АЦП) отслеживать сигнал с фотодетекторов. Так как оптический выходной сигнал записан в смешанном формате, он должен быть преобразован в чисто двоичный код, прежде чем он будет использован в дальнейших расчетах. В целом для двоичных чисел с длиной в l цифр АЦП должен различать l различных уровней. При этом процесс аналого-цифрового преобразования, несомненно, потребляет энергию и может являться самым медленным процессом в цикле обработки сигнала. В целом требования к ширине полосы частот источников и детекторов не являются жесткими. Согласно указанному эмпирическому правилу, если частота передачи данных равна f , то источники и фотодетекторы должны обладать полосой в $4f$. Некоторые из архитектур удовлетворяют требованиям либо к детекторам, либо к источникам (но не обоим одновременно). Архитектуры с пространственным интегрированием дают возможность ис-

¹⁾ В отечественной литературе называемой также теоремой выборки, или теоремой Котельникова. — *Прим. перев.*

пользовать медленное переключение источников из одного состояния в другое со скоростью f/l , в то время как детекторы должны отслеживать все тактовые циклы. Наоборот, архитектуры с временным интегрированием нуждаются в источниках света, переключающихся в каждом тактовом цикле, и детекторах, интегрирующих по l циклам. Всегда могут обнаружиться архитектуры, требующие переключения и источника, и детектора в каждом тактовом цикле.

Влияние полосы пропускания детектора на вычислительные возможности описывается фактором качества, получившим известность как отношение Псалтиса [21]. Отношение Псалтиса представляет собой число операций умножения, приходящихся на одно аналого-цифровое преобразование, которое может выполнять вычислительное устройство. Желательно, чтобы этот параметр имел значение более 1. Указанную величину обычно вычисляют путем деления числа операций умножения в секунду (скорость вычислений) на число аналого-цифровых преобразований в секунду (число выходных каналов, умноженное на полосу пропускания канала). Поскольку скорость вычислений зависит от ширины полосы частот входного сигнала, а ширина полосы частот как входного, так и выходного сигналов связана с общей тактовой частотой f , то отношение Псалтиса не зависит от тактовой частоты. Отсюда вычислительные возможности не могут быть увеличены за счет увеличения частоты работы вычислительного устройства.

Быстродействие оптических матричных умножителей было уже описано в табл. 7.1 и 7.2. При вычислении отношения Псалтиса числа в табл. 7.1 и 7.2 имеют «коэффициент запаса» 2, поскольку были учтены операции умножения и сложения. Для умножителей матриц на векторы соотношение Псалтиса показано в табл. 7.3. Умножители матрицы на матрицы показаны в табл. 7.4. Вычисления проведены для тех же случаев, что и в табл. 7.1 и 7.2. Второй столбец табл. 7.3 предполагает значение $l=16$, $n=m=32$, в то время как для третьего столбца $l=32$, $n=m=128$. В табл. 7.4 второй столбец соответствует значениям $l=16$, $n=m=k=32$, а третий столбец относится к $l=32$, $n=m=k=128$.

Таблицы 7.3 и 7.4 демонстрируют прямо-таки ужасающую неэффективность работы аналого-цифрового преобразователя. По мере увеличения размеров матриц отношение Псалтиса резко падает. Единственной архитектурой, для которой отношение Псалтиса еще как-то приближается к 1, является двумерная архитектура вычисления внешнего произведения. Архитектура этого вида в наибольшей степени использует преимущества двумерной природы распространения света.

Альтернативный подход к проблеме увеличения отношения Псалтиса был предложен в [22]. Авторы этой работы предло-

жили вариант архитектуры с временным интегрированием, в котором операция сдвига выполняется в фотодетекторе, а не во входном регистре. Рассмотрим рис. 7.2: число h вводится в сдвиговый регистр не в виде временной последовательности сигналов, а представлено в виде неподвижной маски (транспаранта) на протяжении всех операций процедуры вычислений. Фотодетектором служит ПЗС-сдвиговый регистр, который смещается и суммирует свое содержимое на каждом шаге. Для обработки входного сигнала требуются буферные разряды, поскольку при точности входного сигнала l цифр результат имеет длину $2l-1$ цифр. Буферы входного сигнала позволяют выделить время на очистку сдвигового регистра от сигнала. Они также позволяют выделить время на загрузку входного модулятора следующим набором цифр. Входной модулятор, используемый в данном случае, аналогичен модуляторам, применяемым в архитектуре вычислений внешнего произведения, в том плане, что цифры должны быть введены за один тактовый цикл, а не загружаться постепенно, как в случае стандартной архитектуры с временным интегрированием.

Для реализации умножения вектора на вектор ряд таких модуляторов соединяют вместе по одному на каждый элемент вектора. Выходные сигналы суммируются на одном детекторе, как в архитектуре 2М ВИ (II). Но в отличие от случая умножителя вектора на матрицу, применяющего комбинацию узлов умножителя вектора на вектор, этот процессор использует один умножитель вектора на матрицу, но пропускает всю матрицу через него синхронно с тактовыми импульсами. Архитектура требует $m(2l-1)$ циклов для умножения матрицы $n \times m$ на вектор $n \times 1$. Как и в случае архитектуры 2М ВИ (II) зависимость от n отражается на требованиях к динамическому диапазону детектора. Соответственно, затраты времени на операцию умножения и рабочие частоты будут такими же, как для случая архитектуры 2М ВИ (II), показанной в табл. 7.1. Однако в данном случае имеется только один выходной канал, в котором аналого-цифровое преобразование должно выполняться в каждом тактовом цикле. В результате (как видно из табл. 7.3), отношение Псалтиса для первой тестовой задачи составляет 1,032 и соответственно 2,032 — для второй.

Умножитель вектора на вектор позволил достичь желаемого отношения Псалтиса за счет двух обстоятельств. Во-первых, использовался модулятор, не требующий последовательной загрузки данных, что позволяет сократить время загрузки. Во-вторых, в последнем случае намеренно ограничено до минимума число оптических операций. Параллелизм обычно рассматривается как главное преимущество оптики, но единственный путь увеличения отношения Псалтиса для оптической системы состоит в уменьшении степени параллелизма.

Сведением этой идеи до абсурда является одноразрядный множитель, через который вся задача пропускается строго последовательным образом. Однако в этом случае множитель становится узким местом, ограничивающим работу всей системы. Урок, который следует извлечь из этого, состоит в том, что для эффективной работы (описываемой отношением Псалтиса), параллелизм должен быть принесен в жертву в уплату за использование аналого-цифрового преобразователя.

Физическая реализация источников и детекторов частот представляет собой более сложную проблему, чем обеспечение требуемой ширины полосы частот. В то время как одномерные и двумерные матричные детекторы стали широко распространенными устройствами уже на протяжении ряда лет, матричные излучатели еще не стали таковыми. Если уменьшить до минимума число детекторов, то тогда архитектура типа 2М ВИ (II), требующая использования матричных источников, выглядит в настоящий момент вполне реализуемой. При этом конечный размер детектирующих элементов и источников также представляет проблему для интерфейсных элементов. В то время как активная область источника, модулятора или детектора может составлять лишь несколько микрон, из-за наличия шин управления активной областью детектор обычно имеет существенно большую площадь. В результате отдельные ячейки фотодетектора обычно имеют размер в несколько десятков микрон.

Из-за расхождения в размерах и наличия определенной пространственной конфигурации дискретные элементы не могут быть скомбинированы в матричные устройства. Межэлементные соединения обычно требуют управления с помощью анаморфных преобразований изображений и оптического преобразования Фурье. Если размеры ячеек составляют порядка 100 мкм, оптика обеспечивает хорошее качество изображения на расстояниях, равных или меньших $f/20$. При уменьшении размеров ячейки или по мере увеличения необходимого усиления, требуемого для согласования компонент, требования к величине f становятся более строгими. Многие архитектуры требуют использования сегментированной оптики или матриц линз. Матрицы голографических линз могут снять остроту некоторых проблем изготовления. Другая возможность, если позволяют размеры ячеек, состоит в использовании волоконной оптики для соединения ячеек модулятора с элементами детектора. (Возможно, курьезом выглядит тот факт, что, для того чтобы воспользоваться преимуществом оптики при реализации большего быстродействия, приходится отказаться от безынтерференционного распространения волн в системе.)

Электрооптические модуляторы представляют собой весьма быстроразвивающуюся область, рассматривая которую трудно избежать устаревания информации. В числе подходов, предло-

женных для реализации различных архитектур, находятся акустооптические ячейки [23], интегрально-оптические дифракционные решетки [24], устройства на поверхностных акустических волнах [25] и магнитооптические модуляторы [26]. Некоторые из архитектур требуют использования еще не существующих видов модуляторов.

Двумерные модуляторы, хотя и не проработаны на высоком уровне, однако очень просты для применения в оптических компьютерах, о чем свидетельствуют параметры архитектуры 2М ВП. В этом случае модуляторы имеют сравнительно большие размеры ячеек (около 100 мкм или более) и ограниченные полосы пространственных частот. Максимальный размер таких модуляторов редко превышает 256×256 адресуемых ячеек. Этот предел имеет важные последствия для архитектур, требующих параллельного ввода в модулятор всех цифр. В частности, отсюда следует, что процессоры, вычисляющие внешнее произведение, исключительно точно «попадают в цель». Например, при точности вычислений в 32 бита (как в табл. 7.2) модулятор может выполнять процедуры только с 8 числами. Если провести разбиение матрицы 128×128 на подматрицы 8×8 , то для решения задачи процессор теперь должен выполнять $16 \times 16 = 256$ последовательных операций умножения матриц. Скорость вычислений уменьшается с 164×10^7 оп/с до $10,2 \times 10^7$ оп/с, что несколько лучше, чем у других умножителей матрицы на матрицу.

В большинстве одномерных модуляторов применяются дифракционные решетки с модулированным профилем, т. е. каждый бит переносится на несущей частоте. Следовательно, согласно эмпирическому правилу, несущая частота должна быть в 4 раза выше, чем тактовая частота. Число цифр, которое может быть размещено в дифракционном модуляторе, эквивалентно тактовой частоте, умноженной на быстродействие ячейки, или $\tau f/4$, где τ — быстродействие ячейки, а f — частота модуляции. Это эквивалентно утверждению, что каждый бит четыре раза представлен несущей частотой. Акустооптические модуляторы обычно имеют произведение τf , достигающее 2000, так что акустооптические модуляторы могут содержать в себе одновременно до 500 цифр. Важно согласовать тактовую частоту с частотой, на которой работает сам модулятор. Если используется тактовая частота в 10 МГц, и для 1-ГГц ячейки $\tau f = 2000$, тогда в ячейке будут содержаться только 20 цифр.

Если модулятор можно модулировать по частоте, то ситуация оказывается несколько более сложной. Модулятор, работающий на частоте f , имеет полосу частот в $f/2$, что вытекает из «теоремы выборки» (сигнал захватывается несущей частотой дважды за цикл). Для простоты далее будем полагать, что в этом диапазоне частот частотная характеристика оказывается

монотонной. Чтобы разделить m строк матрицы, требуется разделить полосу частот на m поддиапазонов с шириной $\Delta f = f/2m$. Глубина модуляции в каждом из поддиапазонов не может превышать Δf . Согласно приведенному выше правилу «четверки», временной интервал, занимаемый одним битом в ячейке, должен составлять не менее чем $\tau_B = 8m/f$, что является обратной величиной к одной четвертой части ширины полосы частот. Если временная апертура ячейки составляет τ , полное число бит, которое может загружаться в ячейку, составляет $\tau/\tau_B = l$, что равняется точности представления числа. Полагая $\tau f/2 = TB$, где TB — время стробирования ячейки, получаем соотношение $4 ml = TB$. Если, как и в предыдущем примере, $\tau f = 2000$, тогда $TB = 1000$ и $ml = 250$. Чтобы сохранить $l = 16$, величина m ограничена значением 15. Для сравнения укажем, что на основе анализа дифракционных процессов авторы [23] получили предельное значение $ml = TB$. Проведенный ими анализ частотных характеристик дает более жесткое ограничение величины — в 4 раза.

7.4. Заключение

В настоящее время, как видно из изложенного выше материала, оптические методы умножения матрицы находятся еще только в стадии зарождения. На текущем уровне работ имеются дискретные источники света, модуляторы и фотодетекторы, которые не слишком просто состыковываются друг с другом и с обслуживающей их электроникой, что ограничило объем решаемых ими задач. Эти устройства выдают выходные сигналы, которые не пригодны для дальнейших вычислений. Можно предположить, что в будущем развитие пространственных модуляторов света, интегральной оптики и микротехнологии приведет к созданию интегрированных оптических систем, включающих все компоненты в один узел.

Гораздо более проблематичным является ограниченная емкость обрабатываемых чисел для всех обсуждавшихся выше архитектур. В начале главы утверждалось, что оптика имеет большую полосу пространственных частот. Действительно, двумерные оптические поля позволяют иметь миллионы разрешающих элементов с линейными размерами в несколько длин волн света. Но управлять регистрацией оптических полей удастся лишь только в масштабах десятков и сотен длин волн. Широкая полоса пространственных частот представляется доступной только для систем формирования изображений. При наличии данного ограничения цифровые оптические процессоры кажутся пригодными лишь для сравнительно ограниченных по объему задач. Кроме того, многие интересующие проблемы (построение калмановского фильтра и даже цифровые преобразования

Фурье) могут быть выполнены в матрицах 128×128 , так что оптические вычислительные устройства могут быть полезны даже в том случае, если объем выполняемых оптических вычислений ограничен.

Самым серьезным недостатком цифровых оптических систем, вероятно, является наличие аналогового выходного сигнала. Оптический сигнал не может быть легко преобразован в цифровой вид. Выходной сигнал должен быть последовательно скорректирован электронными методами при не слишком высоких затратах. Необходимо найти какие-либо способы выполнения преобразования оптическими методами или способы использовать выходной сигнал, представленный в смешанном формате.

Видимо, изложенный способ рассмотрения проблем оптического матричного умножения следует воспринимать лишь как определенный, далеко неокончательный этап развития. За последнее время было получено много новых данных относительно возможностей оптики в вычислениях, а обсуждавшиеся здесь работы простимулировали появление многих приборных разработок. В самом деле, многие из рассматривавшихся в данной главе концепций находят применение во все разрастающейся области оптических межэлементных соединений для нейронных сетей [27]. Применительно к проблеме соединений оптические системы рассматривают как программируемые переключатели, а не вычисляющие устройства.

Выходной сигнал в этом случае нуждается только в пороговом кодировании, а не в аналого-цифровом преобразовании. Избавившись от этих отягощающих проблем, можно использовать в полной мере параллелизм оптических методов.

*Т. Ятагаи, Институт прикладной физики,
Университет г. Цукубэ, Цукубэ, Ибараки, Япония*

8.1. Введение

Бозрастание требований, предъявляемых к скорости обработки больших объемов данных, указывает на необходимость создания компьютеров с параллельной обработкой данных, для которых характерно высокое быстродействие [1, 2]. Для достижения высокой степени параллельности вычислительных систем могут быть использованы методики оптических вычислений. За последнее десятилетие было предпринято много попыток разработать параллельные оптические процессоры [3, 4]. Авторы [5] усовершенствовали схему компьютера Зи (Tse), продемонстрировав практическую значимость и ряд преимуществ оптической параллельной обработки. В качестве устройств параллельной логики использовали оптическую пороговую матрицу [6] и микроканальный усилитель изображений [7]. Многими авторами описывались жидкокристаллические клапанные устройства для модуляции света, в частности устройство Hughes LC использовалось для выполнения различных оптических логических операций и параллельного порогового кодирования [8, 11].

Требуется отчетливо выделить значение различных архитектурных подходов применительно к параллельным оптическим компьютерам, и ряд авторов уже рассматривали этот вопрос с различных точек зрения. В работах [12, 13] описан оптический последовательный процессор, созданный на основе оптических межэлементных соединений, предельные возможности которого обсуждались с позиций анализа числа оптических вентилях. В работах [14, 15] было предложено использовать теневую методику для осуществления оптической цифровой логики, в частности обсуждалось ее применение в оптической системе параллельной логики (OPALS). Авторы [16] применили жидкокристаллический ПМС в оптических триггерах и схемах последовательной логики, а также рассмотрели их применение в архитектуре клеточной логики.

В данной главе дается описание уникальной по своим возможностям и единственной в своем роде архитектуры оптических компьютеров, а именно архитектуры клеточной логики [17, 18].

Можно проследить историю развития клеточной логики до начала 60-х гг. Для реализации обладающих высоким параллелизмом систем обработки предлагалось большое количество видов клеточных матричных структур процессорных элементов. Иногда клеточная логика рассматривалась в связи с теорией нейронных сетей. В настоящее время представляется многообещающим изготовление клеточных матриц на основе технологии СБИС. В свою очередь оптическая параллельная логика подает еще большие надежды потому, что с помощью оптики можно очень просто реализовать пространственное размещение элементов процессора и выполнить соединения между матрицами процессорных элементов. Архитектура клеточной логики в полной мере способствует проявлению таких преимуществ оптических операций, как высокая степень параллелизма. Класс архитектур клеточной логики позволяет эффективно осуществлять функции контроля и управления оптическими параллельными компьютерами. Особой чертой архитектуры клеточной логики является то, что ее математические структуры доступны для понимания во многих случаях. Это — теория клеточной логики, называемая клеточными автоматами. Архитектура клеточной логики и теория клеточных автоматов могут играть важную роль в развитии оптических компьютеров. Главное внимание здесь уделяется подходам, основанным на клеточной логике и рассматриваемых применительно к конструированию оптического компьютера. Особый интерес представляют характеристики и возможности архитектуры клеточной логики и их реализация в виде оптических устройств.

В разд. 8.2 представлен краткий обзор архитектур клеточной логики применительно к разработке компьютера, к чему позднее придется вернуться при описании конкретных разработок. Представлены некоторые клеточные автоматы, а также описаны их наиболее значительные характеристики и результаты. Рассмотрены перестраиваемые архитектуры усовершенствованных клеточных логических компьютеров. Обсуждается классификация методов обработки данных, указано значение МКМД-архитектуры для организации потоков логических данных и команд в клеточных логических компьютерах общего назначения.

В разд. 8.3 даны некоторые основные понятия логики и методы двоичной обработки данных применительно к архитектурам клеточной логики. Затронуты оптические методики реализации МКМД-логики. Показано, что оптическая МКМД-логика позволяет достичь особой гибкости архитектуры оптического компьютера. Наконец, в разд. 8.4 представлены способы реализации архитектуры клеточной логики в оптических устройствах.

8.2. Клеточная логика

Архитектура фон Неймана и теория автоматов легли в основу разработки электронных цифровых компьютерных систем [19]. Однако в случае компьютеров с чисто параллельной обработкой данные принципы неприменимы. Было показано, что эффективное решение в случае чисто параллельной архитектуры имеется лишь при определенных условиях. Клеточная логика среди различных архитектур [20—22] является одним из наиболее вероятных кандидатов на эту роль. Архитектура клеточной логики для оптических компьютеров основана на использовании упорядоченных простых процессорных элементов, или элементарных блоков логических операций. В целом реализация клеточной логики — это пространственное расположение ячеек процессорных элементов в одном, двух или трех измерениях. В принципе размещение должно быть до некоторой степени унифицировано, однако в соответствии с конкретной ситуацией может изменяться. Каждая ячейка в клеточной матрице обладает определенными логическими свойствами и может также обладать способностью запоминать информацию. Клеточная матрица характеризуется однородным распределением соединений между ячейками.

В настоящее время ведется интенсивная разработка по следующим направлениям: теория клеточных матриц, теория клеточных автоматов, клеточные логические компьютеры, возможные операции и функции в клеточной логике, свойства клеточной логики. Важной задачей в исследованиях клеточной логики является получение таких схем переключения и хранения информации, которые бы были лучше, дешевле, меньше по размеру и надежнее, чем существующие схемы и компьютерные системы. Важные области исследований клеточной логики включают в себя нижеследующее:

1. Синтез логических функций.
2. Клеточные логические автоматы.
3. Клеточные логические компьютеры.

Вопрос о синтезе клеточных логических функций обсуждается в связи с его важностью для управления оптическими компьютерами. Далее рассматривается вопрос об оптической реализации клеточных логических компьютеров, основанных на клеточных автоматах. В некоторых клеточных системах логики параллельно выполняются различные типы операций. Это означает, что оптические методики пространственно-инвариантного фильтрации могут быть применены к локальным клеточным логическим операциям способом, аналогичным тому, как это делается в традиционных оптических вычислениях, основанных на оптическом преобразовании Фурье. Архитектуры локальной клеточной логики описаны в разд. 8.4.2.

Таблица 8.1. Архитектуры компьютеров, используемые для обработки изображений

Архитектура клеточной логики	Чисто параллельная схема	CLIP-4 (Лондонский университет)
		MPP (NASA) ILLIAC IV
Архитектура конвейерной обработки	Локально параллельная схема	Tospix (Тошиба, Япония)
		Comtal (De Anza) DIP (Дельфтский университет)
		IU-проект Систолическая матрица (CMU) WOWS (Фэйрчайльд корп.)

8.2.1. Клеточные компьютеры

До настоящего момента для цифровой обработки изображений было разработано много различных типов электронных клеточных компьютеров [23—25]. Как показано в табл. 8.1, клеточные логические компьютеры подразделяются на две категории:

1. Чисто параллельные.
2. Локально параллельные.

На рис. 8.1 показан типичный клеточный логический компьютер, основанный на чисто параллельной схеме. Главная часть клеточного компьютера — это матрица процессорных элементов (ПЭ). Обычно всеми ПЭ управляют с помощью одной шины команд. Каждый ПЭ независимо выполняет одну и ту же операцию. Другим классом клеточной логики является локальная клеточная логика. На рис. 8.2 изображено формирование локально параллельной клеточной логики. Например, входное подлежащее обработке изображение состоит из $M \times M$ ячеек. Операции локальной клеточной логики выполняются только для малых $N \times N$ элементов матриц с целью получения одного выходного элемента, где $N \leq M$. Обычно все входные элементы обрабатываются одним и тем же типом операций. Параллельно обрабатываются только маленькие площади входных элементов. При этом фрагмент входного сигнала последовательно сканирует по входной матрице данных. Возможно, наиболее очевидным применением такого оптического логического клеточного компьютера является распознавание образов, поскольку оно относится к задачам обработки изображений.

Локальной логической функцией окрестности называется такая функция, в которой величина выходного сигнала в каждом элементе является функцией значения параметра самого элемента и значений ближайших соседних элементов. Процедуры клеточной логики могут быть использованы для нахождения среднего, дифференцирования, нелинейного усиления изображе-

ния, обнаружения края, анализа текстуры, уменьшения толщины и выделения остова. Почти все эти операции, использовавшиеся в цифровой обработке изображений, могут выполняться этим типом компьютера.

Существуют четыре типа локальных клеточных логических компьютеров: (1) параллельный полутонный, (2) последовательный полутонный, (3) параллельный двоичный и (4) последовательный двоичный. Как указано выше, одни и те же операции выполняются для каждой входной матрицы элементов во многих клеточных логических операциях. Используя термины из области оптической фильтрации, этот тип операций клеточной логики является пространственно-инвариантным. Это предполагает, что традиционные методики пространственно-инвариантной оптической фильтрации могут быть применены при реализации операций клеточной логики для параллельного полутонного и параллельного двоичного форматов данных.

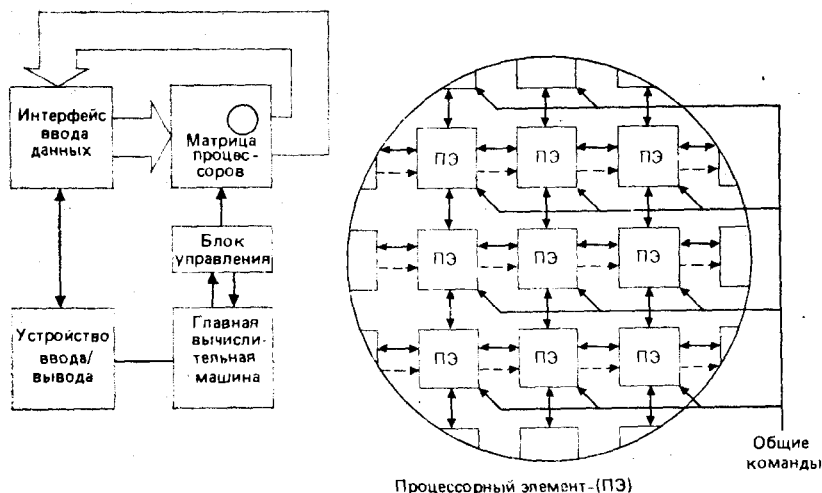


Рис. 8.1. Пример клеточного логического компьютера с чисто параллельной архитектурой.

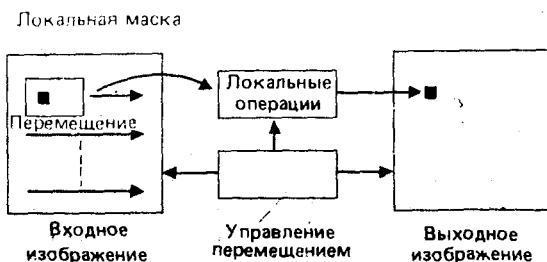


Рис. 8.2. Пример компьютера с локальной клеточной логикой.

8.2.2. Структуры матриц

Было предложено много видов клеточных матриц. Структура клеточной матрицы существенно влияет на функции и особенности системы. На рис. 8.3 показаны примеры различных типов клеточных матриц, включая прямоугольную, гексагональную и треугольную. Здесь каждый ПЭ может непосредственно соединяться с соседними элементами. Обычно для простоты полагают, что каждая ячейка соединена с ее четырьмя соседями по горизонтали и вертикали. Теоретически структуры клеточной матрицы не являются обязательно двумерными. Например, тороидальная и гиперкубическая матрицы могут быть представлены как двумерные сети. Выбор типа матрицы осуществляется на основе компромисса между возможностью реализовать соединения и требуемыми характеристиками.

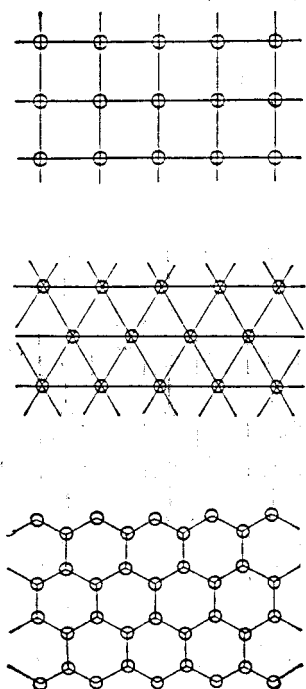


Рис. 8.3. Структуры клеточных матриц.

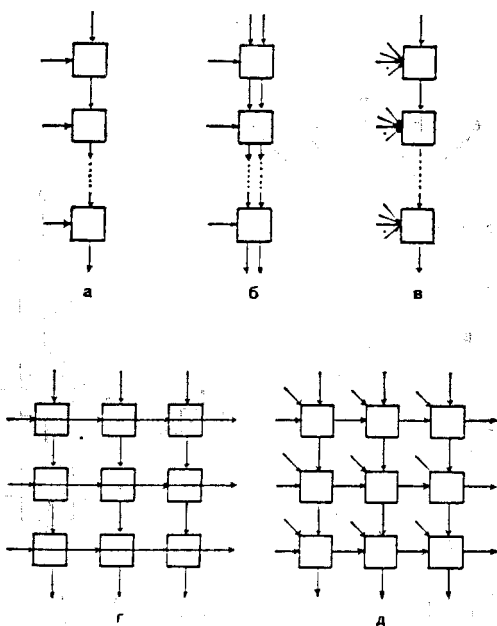


Рис. 8.4. Варианты размещения элементов в схеме клеточной логики: а — каскад Мейтры; б — двухканальный каскад; в — каскад с большим числом входов; г — матрица Минника; д — матрица мажоритарности.

8.2.3. Клеточный автомат

Для управления чисто параллельным оптическим компьютером важное значение имеют комбинации произвольных логических функций. На рис. 8.4 показаны типичные структуры размещения ячеек, предложенные для синтеза логических функций. Проблемы одномерного логического размещения ячеек были решены в [26]. Такие сети в виде цепей называются каскадами Мэйтри, или приточными сетями [27]. Возможность существования логических функций и методов их синтеза вполне ясна. В двухканальном каскаде одна выходная линия подсоединяется к каскадам Мэйтри, так что может быть синтезирована произвольная булева функция [28]. Двумерные расширения клеточной логики описаны в [29—31]. В следующем разделе рассматривается клеточная матрица Минника [29, 30]. Она представляет собой простую и ясную клеточную структуру.

8.2.4. Клеточная матрица Минника и синтез логических функций

На рис. 8.5, б приведен пример клеточной матрицы Минника, а на рис. 8.5, а даны обозначения символов ячейки. Горизонтальный выходной канал x просто соединяется с шиной с гори-

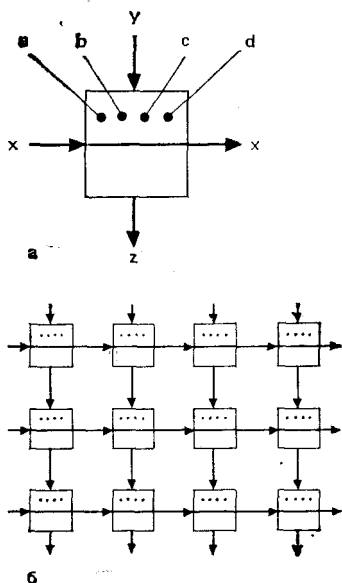


Рис. 8.5. Клеточная матрица Минника: а — схема отдельной ячейки; б — базовая матрица ячеек Минника.

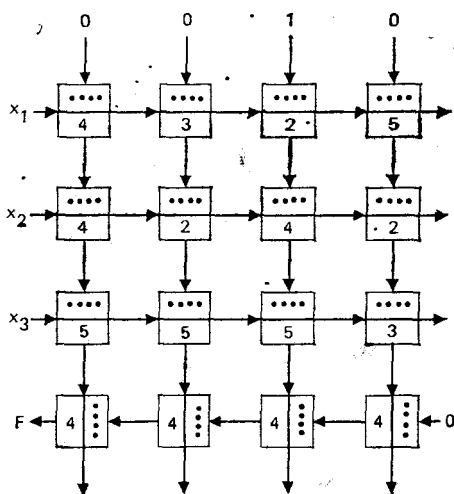


Рис. 8.6. Клеточная матрица Минника для случая трех переменных.

Таблица 8.2. Функции, реализуемые ячейкой Минника «с переключающими точками»

Номер	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>z</i>
0	0	0	0	0	1
1	0	0	0	1	y'
2	0	0	1	0	$x' + y'$
3	0	0	1	1	$x'y'$
4	0	1	0	0	$x + y$
5	0	1	0	1	xy'
6	0	1	1	0	$x \oplus y$
7	0	1	1	1	0
13	1	1	0	1	$x = S, y = R$

горизонтальным входным каналом x . Управление сигналом, появляющимся на нижнем выходе z и подаваемым на верхний вход y соседней ячейки, осуществляется с помощью четырех управляющих каналов, обозначенных точками (табл. 8.2). Имеется только девять разрешенных состояний «переключающих точек» управляющих шин: 0, 1, шесть логических функций и состояние триггера (с включенным и выключенным состояниями R и S). Одним очень важным результатом идей Минника является то, что при подаче на «переключающие точки» сигналов, указанных в табл. 8.2, могут быть реализованы логические функции, представленные в столбце б. Такой теоретический подход (или теория клеточного автомата) является существенно необходимым для разработки оптического компьютера с высокой степенью параллельности.

Ячейка Минника с «переключающимися точками» в основном способна воспроизвести произвольную логическую функцию двух переменных. Фактически все логические функции могут быть получены в результате комбинации ячеек Минника. Например, произвольная функция четырех переменных может быть разложена на четыре произвольные логические функции двух переменных. Минник показал возможность создания декодеров, схем прибавления единицы, сдвиговых регистров, двоичных сумматоров-вычитателей и т. д. с помощью ячеек с «переключающимися точками».

На рис. 8.6 показан пример клеточной матрицы Минника для следующей логической функции:

$$F(x_1, x_2, x_3) = x'_1 x'_2 x_3 + x'_1 x_2 x_3 + x_1 x'_2 x_3 + x_1 x_2 x'_3, \quad (8.1)$$

где x' обозначает логическое отрицание x . Первый член $x'_1 x'_2 x_3$ в уравнении (8.1) получается в первых трех ячейках первого столбца. Второй член $x'_1 x_2 x_3$, третий $x_1 x'_2 x_3$ и четвертый $x_1 x_2 x'_3$

члены получаются во втором, третьем и четвертом столбцах соответственно. В последней строке для всех четырех членов вычисляется логическое ИЛИ.

8.2.5. Перестраиваемые матрицы

Если имеется возможность изменять межэлементные соединения клеточной матрицы Минника, показанной на рис. 8.6, то можно сконструировать перестраиваемую или реструктурируемую клеточную ячейку [32, 33]. Свойство перестраиваемости клеточных матриц обеспечивает возможность программируемости таких вычислительных систем. Располагая перестраиваемой системой соединений, можно приступить к реализации идеи самовоспроизводящихся автоматов, способных копировать их собственную сеть соединений [34]. В оптических системах клеточной логики внесение таких изменений в межэлементные соединения осуществляется сравнительно просто по сравнению с электронными системами.

8.2.6. Классификация методов обработки данных

В соответствии с различиями в потоках данных, потоках команд, а также в степени параллелизма в [35] была предложена классификация способов обработки по следующим четырем основным методам обработки:

ОКОД — с одним потоком команд и одним потоком данных,

ОКМД — с одним потоком команд и многими потоками данных,

МКОД — с многими потоками команд и одним потоком данных,

МКМД — с многими потоками команд и многими потоками данных.

На рис. 8.7 показаны различия между четырьмя классами методов обработки данных. Традиционный цифровой компьютер с архитектурой фон Неймана основан на методе ОКОД. В данном виде компьютеров потоком команд управляет один программный счетчик, а выборка данных осуществляется согласно адресу, содержащемуся в командах. В некоторых клеточных логических компьютерах для цифровой обработки изображений принят метод ОКМД, в котором все элементы процессора начинают одну и ту же операцию по одной команде из управляющего компьютера, как показано на рис. 8.1. Компьютер с конвейерной архитектурой относится к классу МКОД. В МКМД независимые команды с потоками данных выполняются параллельно. Метод ОКМД обладает низшей степенью параллелизма, в то время как МКМД — наивысшей. Архитектуры ОКМД,

и МКМД подходят для параллельной обработки упорядоченных данных, таких как матрицы или изображения. Подход локальной клеточной логики представляет собой архитектуру ОКМД и рассматривается как операции с соседними элементами. Для оптических компьютеров общего назначения архитектура МКМД превосходит ОКМД с точки зрения гибкости и возможности проектировать компьютерные системы.

8.3. Оптическая двоичная логика и архитектура МКМД

Методы оптической параллельной обработки, используемые для выполнения двоичных логических операций, являются ключевыми элементами развивающихся оптических компьютеров. Было предпринято много попыток достигнуть увеличенной пропускной способности двоичной цифровой логики путем использования принципа параллельной обработки, реализуемого оптическими методами.

Авторы [14, 15] описали параллельный оптический логический матричный процессор, построенный на теневой системе и светодиодах. Все 16 логических функций для двух двоичных переменных осуществляются параллельно путем изменения заливки, создаваемой с помощью светодиодов. В работах [36, 37] был предложен оптический логический процессор, работающий на принципе пространственной фильтрации. В двух выше-

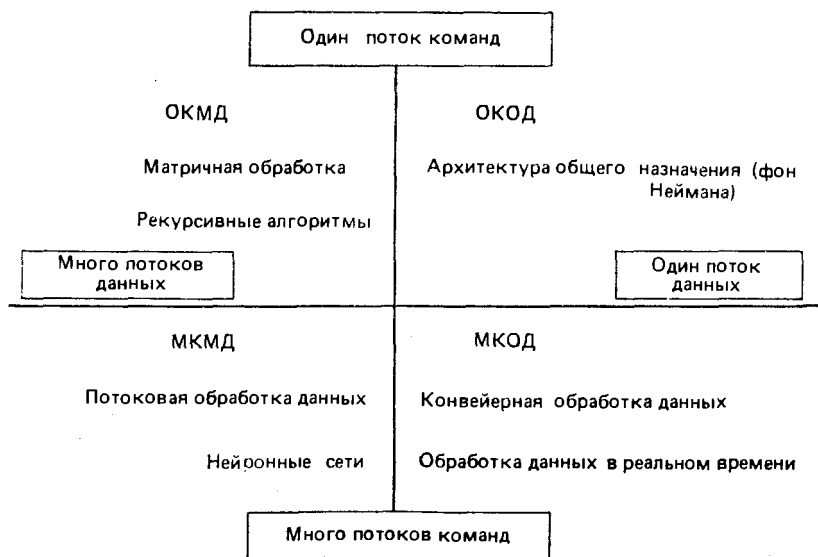


Рис. 8.7. Потоки команд и потоки данных в архитектурах обработки информации.

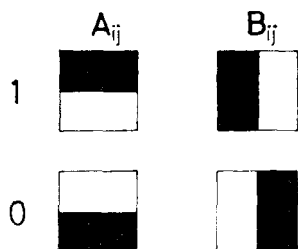


Рис. 8.8. Принципы кодирования в квадратной ячейке.

указанных методах два входных двоичных изображения накладываются друг на друга; результат высвечивается светодиодами или выделяется пространственной фильтрацией. Основная черта обоих методов состоит в том, что одна логическая функция воздействует параллельно на все элементы входного сигнала. Это означает, что описанные выше методы относятся к методу ОКМД.

Здесь описывается другой тип оптической двоичной логики, в котором логические операции в различных точках пространства могут выполняться параллельно [38]. Это обеспечивает оптическую реализацию архитектуры МКМД.

Как в теновом методе, так и в методе пространственной фильтрации, двоичное входное изображение пространственно закодировано. Данные двоичного входного изображения A и B разделяют на $N \times N$ квадратных ячеек. Чтобы реализовать со-

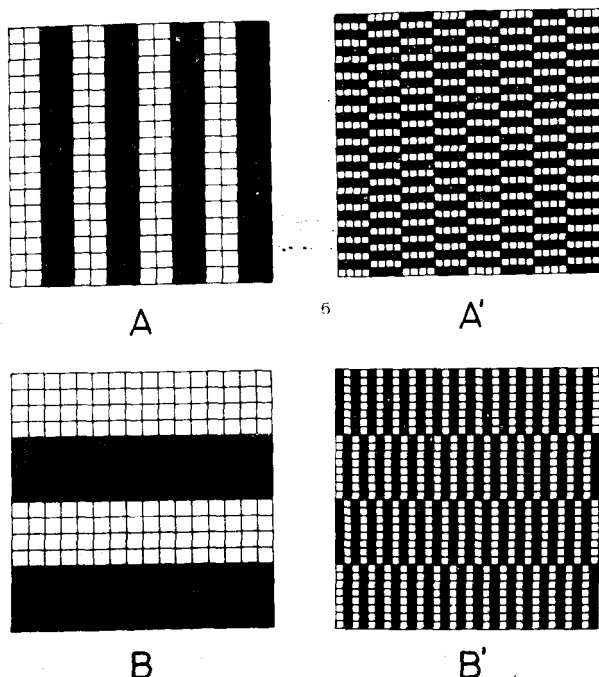


Рис. 8.9. Черно-белое кодирование двух входных изображений: a — два входных изображения A и B ; b — их закодированные двоичные образцы A' и B' .

стояние логических 1 и 0, каждую квадратную ячейку делят на две подъячейки. Принципы кодирования квадратных ячеек показаны на рис. 8.8. Ячейки одного из входных двоичных изображений, обозначенные A , кодируются в горизонтальном направлении, в то время как ячейки другого входного двоичного образца, обозначенные B , кодируются в вертикальном направлении.

Существует много методов кодирования подъячеек, использующих коэффициенты пропускания, поляризацию и рассеяние. Когда применяют методы кодирования черное — белое или прозрачное — непрозрачное, входная ячейка разделяется на изображения, составленные из черно-белых прямоугольников. Каждая ячейка входного изображения A разделена по горизонтали на два прямоугольника, которые закодированы черным цветом (непрозрачный) или белым (прозрачный) в соответствии со своим значением. Другими словами, если ячейка соответствует логической единице, верхняя подъячейка является черной, или непрозрачной, а нижняя подъячейка — белой, или прозрачной. Другое входное изображение B кодируется аналогичным образом, но оно разделено на вертикальные прямоугольники. На рис. 8.9 приведены примеры входных изображений, состоящих из 16×16 квадратных ячеек, и их закодированные двоичные образцы A' и B' . Два пространственно закодированных двоичных образца наложены друг на друга, в результате чего получен закодированный логический образ, показанный на рис. 8.10.

Для того чтобы реализовать двоичную логику, используют особую декодирующую маску, ячейки которой разделены на четыре подъячейки. Расположение ячеек в декодирующей маске выбирается таким же, как у входных изображений. Структуры ячеек декодирующей маски, выполняющие 16 логических функций двух двоичных переменных, показаны на рис. 8.11. Декодирующие ячейки размещаются таким образом, что необходимые логические операции могут быть выполнены в необходимых местах входного изображения. Закодированные логические образы наблюдают через декодирующую маску. Пример декодирующей маски показан на рис. 8.12. В верхней правой части выходного образа осуществляется логическая операция A ИЛИ B , в то время как логическое A осуществляется в нижней левой части. Так как входные изображения на рис. 8.9 имеют 16×16 квадратных ячеек, их декодирующая маска состоит из 32×32 подъячеек.

На рис. 8.13, *a* показан результат осуществления логических операций, выполняемых путем наложения закодированного логического изображения, показанного на рис. 8.10, и декодирующего изображения, показанного на рис. 8.12. Если один из выходных сигналов подъячеек в каждой ячейке является белым (или прозрачным) прямоугольником, тогда выходной сигнал

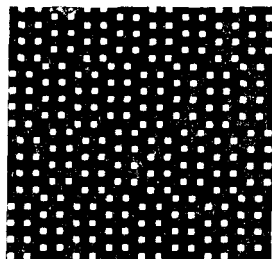


Рис. 8.10. Результат наложения двух закодированных входных изображений.

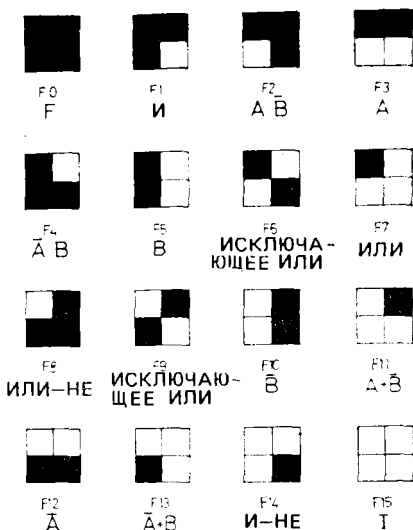


Рис. 8.11. Структура ячеек декодирующих масок для 16-ти логических функций.

ячейки есть логическая 1. Другими словами, выходной сигнал ячейки описывается функцией логическое ИЛИ, реализуемой для выходных сигналов с подъячеек. Для сравнения восстановленное изображение для той же самой логической операции, для тех же самых входных изображений, приведено на рис. 8.13, б.

Выше была описана методика выполнения оптических логических операций, основанная на пространственном кодировании и суперпозиции декодирующей маски с закодированными входными изображениями. С ее помощью могут быть реализованы шестнадцать логических функций двух переменных. В пред-

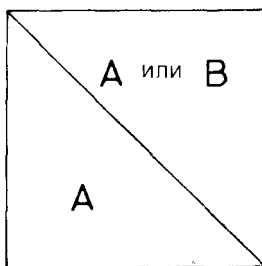
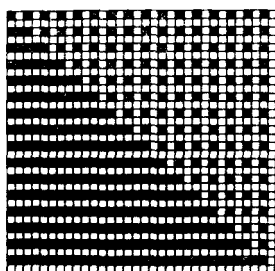


Рис. 8.12. Пример декодирующей маски. Логическая операция A ИЛИ B осуществляется в верхней правой части, логическое A осуществляется в нижней левой части,

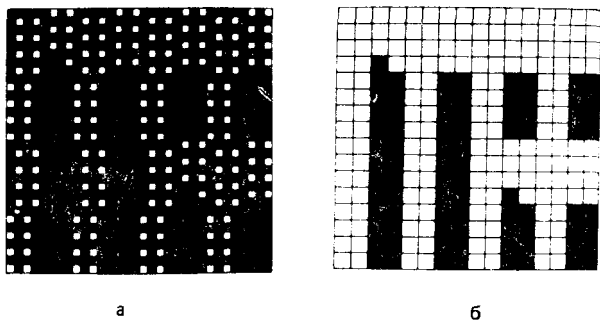


Рис. 8.13. Результат логических операций, выполняемых с помощью декодирующей маски на рис. 8.12: *a* — результат наложения декодирующей маски на рис. 8.12 и закодированное входное изображение $A'B'$ на рис. 8.10; *b* — восстановленный вариант изображения той же маски.

ставленном методе различные логические операции выполняются параллельно. Следует заметить, что в [12] обсуждалась система оптической последовательно реализуемой логики на основе оптической матрицы вентилях ИЛИ-НЕ, а в [19] был предложен ряд оптических логических схем, включающих в себя вентили ИСКЛЮЧАЮЩЕЕ ИЛИ и И. Однако достоинством описанной выше системы является то, что могут быть использованы самые различные логические вентили. Это уникальная особенность характерна для МКМД-архитектуры обработки информации. Данная архитектура предполагает, что представленный метод позволяет уменьшать необходимое число оптических вентилях и расширяет возможности конструирования оптических вычислительных систем. Данный тип двоичной оптической логики мог бы быть использован в оптических компьютерах общего назначения, включая оптические компьютеры с клеточной логикой.

8.4. Оптическая реализация

8.4.1. Оптическая клеточная матрица Минника

На рис. 8.14 показана функциональная блочная схема оптической реализации клеточной логической матрицы Минника. Входное изображение записано в ПМС₁ и подвергается обработке методами параллельной оптической логики. Параллельные логические операции выполняются различными способами. Оптический теневой метод [14] является одним из них. Поскольку эта методика использует оптическое проецирование изображений, то эту методику обычно используют для реализации операций ОКМД, т. е. одни и те же логические операции выполняются для всех элементов входного изображения. Теневой ме-

тод, принципы которого были описаны в предыдущем разделе, может выполнять операции, присущие для МКМД. Выходное изображение в системе параллельной логики записывается в ПМС₂. Оттуда оно отображается на плоскость изображений 1 и, если необходимо, на плоскость изображений 2.

Когда ячейки Минника соединены между собой, выходная информация от декодированного изображения направляется в устройство памяти для хранения изображений, имеющее структуру соединений, показанную на рис. 8.5, б. В логике Минника сети межэлементных соединений являются простыми и упорядоченными, и обычно их основная часть состоит только из устройств сдвига в направлении x и y . Сети соединений реализуют с помощью пространственно-зависимых голограмм [4]. Второе выходное изображение с произвольной временной задержкой подается на входное изображение в качестве сигнала обратной связи, что осуществляется с помощью комбинации пространственного модулятора света и оптического затвора.

В клеточной матрице Минника МКМД тип логики не всегда является необходимым. Однако использование этой логики обеспечивает во многих случаях простую структуру оптических

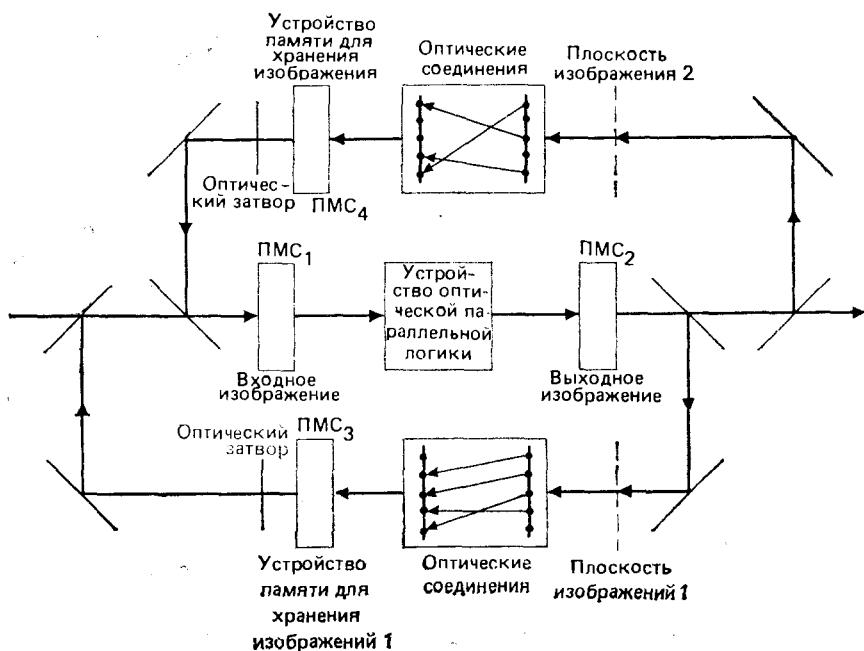


Рис. 8.14. Оптическая схема, реализующая логическую клеточную матрицу Минника.

соединений. В примере, представленном на рис. 8.6, для входных x_i могут быть использованы упорядоченные горизонтальные соединения, потому что в каждой строке к следующей ячейке передается одно и то же значение. Для y_i результаты выполненных операций передаются по вертикальным соединениям. Как показано на рис. 8.15, эти соединения могут осуществляться с помощью цилиндрических линз и за счет наклона зеркала обратной связи.

8.4.2. Локальная клеточная логика

8.4.2.1. Голографическая клеточная логика

В некоторых типах параллельных полутоновых клеточных компьютеров имеется сходство схемы обработки данных с процедурой голографической пространственной фильтрации. Поскольку одна и та же процедура оптической фильтрации применяется к входному изображению, то этот класс операций рассматривается как относящийся к ОКМД. Некоторые операции данного класса основаны на вычислении интеграла свертки, легко осуществляемом оптическими методами. Соответственно оптический клеточный логический компьютер может быть сконструирован на основе голографической пространственной фильтрации.

Рассмотрим операции над матрицами 3×3 элементов. Существует много типов матричных операций; например, для обнаружения градиентов были предложены оператор Приюитта [39]

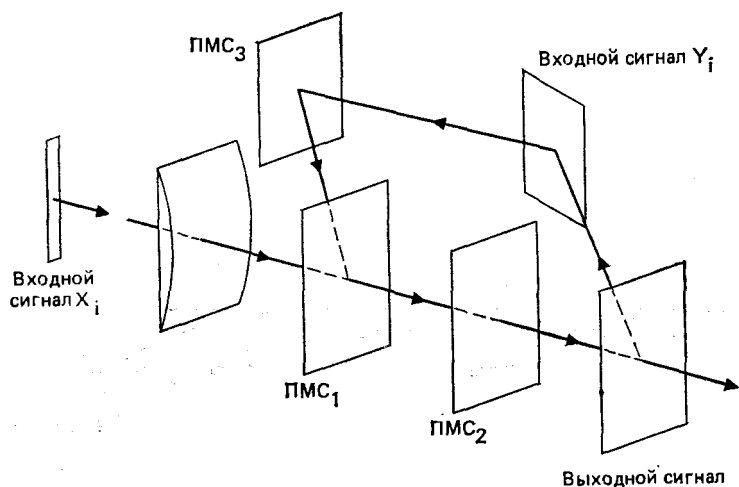


Рис. 8.15. Схема оптических межэлементных соединений, использующая цилиндрические линзы и наклонное зеркало обратной связи,

и оператор Собеля [40]. В методе обнаружения оптимально согласованной границы, предложенном Киршем [41], восемь вариантов операторов действуют на входные изображения и модель, соответствующая максимальному выходному сигналу, определяется как оптимальная модель границы. Эти операции с соседними элементами могут быть описаны процедурой свертки и таким образом являются хорошим примером локальной клеточной логики, реализуемой голографическими пространственно-инвариантными методиками. В [42] была предложена аналогичная система для фильтрации по методу Собеля, в которой фильтр формировался методом многократной экспозиции.

На рис. 8.16 показана оптическая схема, соответствующая предложенной выше системе и состоящая из системы синтеза голографического фильтра и системы фильтрации. С помощью матрицы оптических затворов со сложной амплитудной модуляцией генерируется импульсный отклик в виде матрицы размером 3×3 элемента. ПМС₁ записывает в виде голограммы сложный фильтр для осуществления корреляции с входным изображением. Поскольку сложный фильтр может изменяться в реальном времени, то эта голографически реализуемая клеточная логика является программируемой. В системе фильтрации (вдоль горизонтальной оптической оси) входное изображение записывается в ПМС₂. В представленной системе одновременно используются два лазера с различными длинами волн, применяемые для синтеза фильтров и фильтрации. Таким образом размер ячейки матрицы операций должен быть подобран, чтобы компенсировать увеличение ошибки вследствие наличия двух различных длин волн.

Для проверки принципа голографической клеточной логики был проведен простой эксперимент, в котором использовалась система, состоящая из отдельных подсистем синтеза фильтра

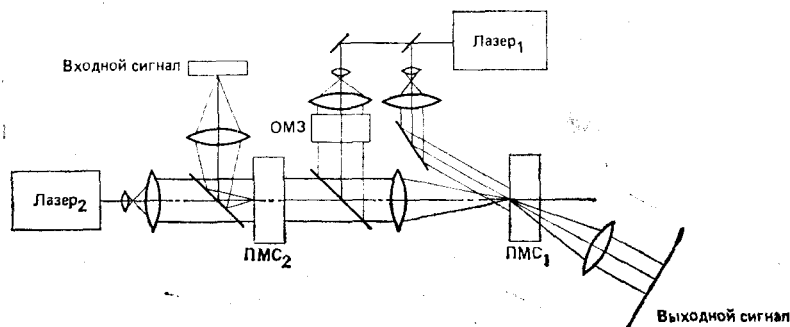


Рис. 8.16. Голографическая система обработки данных с локальной клеточной логикой: ОМЗ — оптический матричный затвор, производящий сложную амплитудную модуляцию; ПМС — пространственный модулятор света.

и системы фильтрации. Было продемонстрировано выполнение 80×80 параллельных операций, реализуемых с помощью голографии [18].

8.4.2.2. Алгоритм перекодированных таблиц и согласованная фильтрация

Архитектура перекодировочных таблиц является мощным классом операций в рамках гибкой двоичной клеточной логики [43]. В таблицу вносятся все возможные выходные сигналы для всех возможных входных сигналов, и работа этой перекодировочной таблицы основывается на соответствующем выборе выходного сигнала для заданного входного сигнала. Для операций в клеточной логике над матрицами малого размера ожидается использование таблиц приемлемого размера. Предположим, например, что размер матрицы клеточной логической операции составляет 3×3 элемента. Содержимое 3×3 элементов матрицы преобразуется в 9-разрядное двоичное число, которое используется в качестве адреса для перекодировочной таблицы, состоящей из 512 элементов. Указанный процесс изображен на рис. 8.17. За счет изменения перекодировочной таблицы могут быть реализованы разнообразные двоичные операции [44].

Процесс табличного поиска является разновидностью операции сопоставления с образцом. Такой тип обращения к таблице или памяти называют адресуемой к содержанию, или ассоциативной, памятью. Оптическая ассоциативная память может быть реализована на основе методики оптической согласованной фильтрации [45] и методики оптических символьных подстановок [46]. В методе символьных подстановок используют параллельное распознавание образа и замену его на другой образ.

В оптических системах табличный поиск может быть проведен в обратном порядке по сравнению с упомянутыми выше случаями электронных систем. То есть ведется поиск изображения размером 3×3 элемента, имеющего необходимое содержание. Если содержимое перекодировочных таблиц является нулевым, тогда на выходной сигнал накладывается шаблон и/или ведется поиск образов, чьи выходные сигналы являются логическими функциями.

Методика оптической согласованной фильтрации может быть применена для поиска или согласования нужных выходных образов. Однако процедура сопоставления с образцом не является одинаковой для всех элементов и, следовательно, этот тип операций, используя терминологию из области оптической фильтрации, называют пространственно-зависимым.

На рис. 8.18 показана схема поиска в перекодировочных таблицах с помощью согласованных фильтров. С помощью со-

гласованного фильтра определяется положение образа с определенным содержанием, чей выходной сигнал должен быть логической 1. Эта процедура повторяется до тех пор, пока не найдены все образы с выходным сигналом 1. Результаты согласованной фильтрации накладываются друг на друга, чтобы получить окончательное выходное изображение. На рис. 8.18 для изменения согласованных фильтров используется последовательно сдвигаемая линейная матрица согласованных фильтров. Акустооптические устройства могут быть использованы для

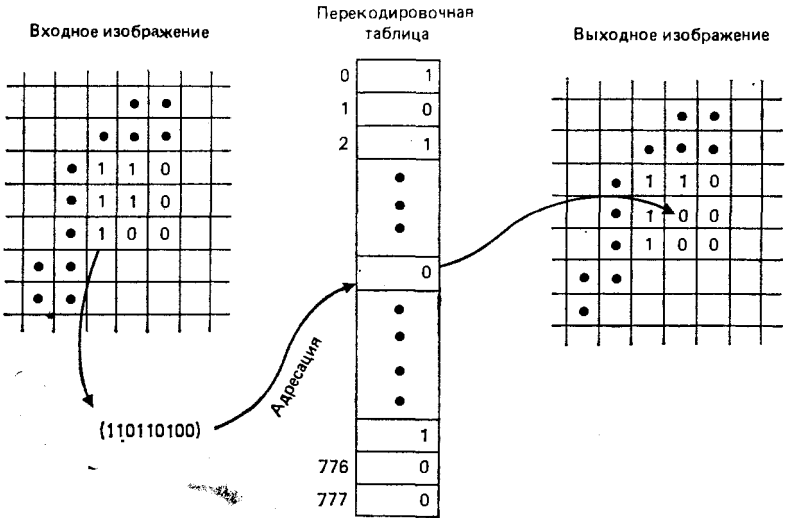


Рис. 8.17. Архитектура перекодировочной таблицы.

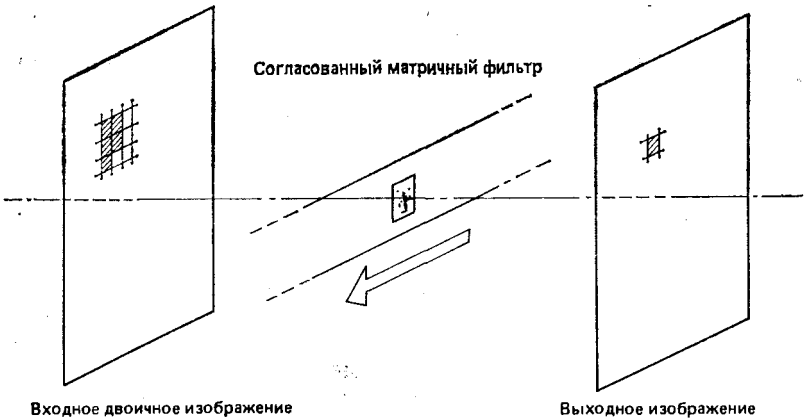


Рис. 8.18. Определение перекодировочной таблицы с помощью согласованного фильтра.

быстрого перемещения матрицы согласованных фильтров. Наложение выходных изображений, полученных от ряда согласованных фильтров, дает окончательный результат.

8.4.2.3. Клеточная логика, использующая матрицы согласованных фильтров

В схеме, показанной на рис. 8.18, процедура поиска каждого желаемого элемента образа является параллельной, но обработка входного сигнала во времени в целом осуществляется последовательно. Параллельная процедура обработки может выполняться с помощью системы мультиплексирования изображения, включающей матрицу микролинз и матрицу согласованных фильтров. На рис. 8.19 показан пример оптической схемы, основанной на методике согласованной фильтрации. В случае операций с матрицей 3×3 элемента требуется перекодировочная таблица с 512 элементами, а поиск образов, соответствующих 512 элементам, содержащимся в таблице, ведется с помощью 512 различных согласованных фильтров. Для того чтобы реализовать параллельную обработку данных, используются многоканальные системы, обрабатывающие большое число изображений, а параллельная согласованная фильтрация выполняется с помощью матрицы микролинз. Для выполнения многоканальной обработки изображений может быть использована высокоэффективная волоконная решетка, составленная из монослойных оптических волокон [47]. Для реализации параллельной согласованной фильтрации используется матрица микролинз с градиентным распределением показателя преломления [48]. Сзади от матрицы согласованных фильтров расположена перекодировочная двоичная таблица. Функция, выполняемая системой, может быть изменена путем замены оптической маски, соответствующей перекодировочной таблице. Результат наложе-

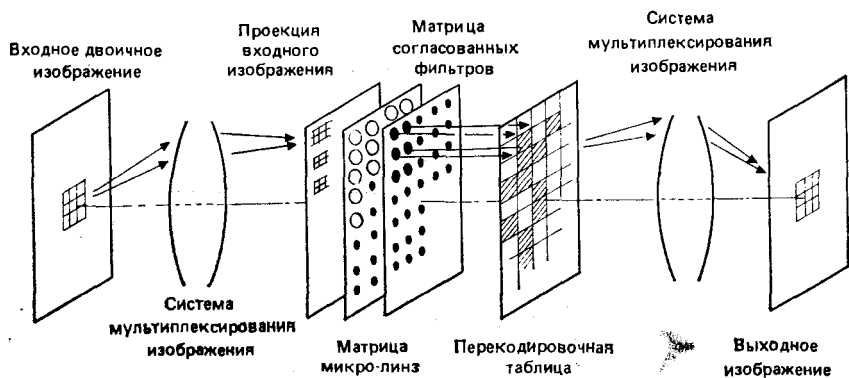


Рис. 8.19. Схема оптической клеточной логики, использующая матрицу согласованных фильтров,

ния образа, создаваемого перекодировочной таблицей и выходного сигнала матрицы согласованных фильтров, является образом, демультимплексированным в выходной плоскости, и представляет собой окончательный выходной сигнал системы.

Была сконструирована более простая схема, в которой используется матрица полупроводниковых лазеров [49], освещающих систему так, как показано на рис. 8.20. Матрица входных изображений с малыми размерами освещается матрицей полупроводниковых лазеров. Если лазеры некогерентны, то фильтрация разных элементов может выполняться одновременно, обеспечивая тем самым параллельную обработку.

8.5. Заключение

В данной главе обсуждалось значение архитектур клеточной логики для оптических компьютеров. Предложены некоторые схемы клеточных логических систем с целью проработки конструкции оптических компьютеров общего назначения.

Клеточная матрица Минника является одним из наиболее мощных вариантов оптических клеточных матриц. Показано, что оптические МКМД варианты схем имеют большое значение для реализации архитектур клеточной логики. Использование МКМД-логики обеспечивает возможность создания простых схем межэлементных соединений.

Те или иные варианты оптической клеточной логики могут обеспечить создание перестраиваемых и самовоспроизводящихся автоматов, основанных на динамических схемах соединений. Наконец, современные тенденции показывают, что область клеточной логики, согласно ожиданиям, значительно обогатится в связи с развитием теории нейронных сетей.

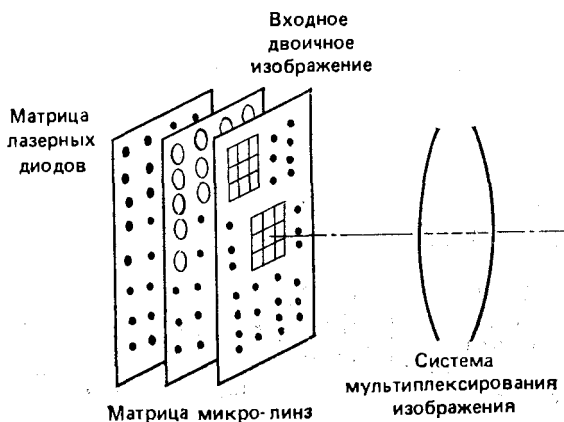


Рис. 8.20. Схема мультиплексирования изображения с помощью матрицы лазерных диодов.

*Р. Арратун, Отделение электро- и вычислительной
техники, Государственный университет им. Уэйна,
Детройт, шт. Мичиган*

9.1. Введение

На протяжении последнего десятилетия развитие оптических вычислений было ограничено преимущественно системами, основывающимися на аналоговой обработке [1]. Достижения цифровой оптической обработки оказались сравнительно слабыми, отчасти из-за того, что оптика «слишком» хорошо подходила для параллельных аналоговых операций, и отчасти из-за принципиальных трудностей, связанных с рассеянием мощности в оптических переключающих элементах. Часть ограничений, связанных с рассеянием тепла для оптических переключающих устройств, была исследована в [2]. В более поздней работе [3] автор детально исследовал этот вопрос и количественно описал те или иные достоинства широкого круга электронных и оптических переключающих элементов. Автор [3] пришел к выводу, что, за исключением очень больших скоростей переключения, оптическая логика не дает особенных преимуществ по сравнению с электронными логическими схемами. Его результаты демонстрируются на рис. 9.1, где представлены параметры, характеризующие энергию, мощность и полосу частот разнообразных электронных и оптических переключающих элементов. Когда рассматривается вопрос об относительных размерах устройства, в большинстве случаев сравнение характеристик приводит к выводу, что, за исключением наиболее специфических областей применения, возможности оптических логических устройств невелики. Одной из таких областей являются системы оптической связи. Если носителем информации является сам световой пучок, тогда применение оптических модуляций и переключения является естественным и удобным. В отличие от переключающих устройств устройства оптической связи уже сейчас используются в существующих компьютерных системах для реализации сложных схем соединений на уровнях плата — плата и чип — чип. Согласно принятому подходу, в данной главе рассматриваются попытки выполнить чисто комбинаторные логические операции на внутричиповом уровне с помощью электроники или реализовать переключающие элементы оптоэлектронными методами, а межэлементные соединения — опти-

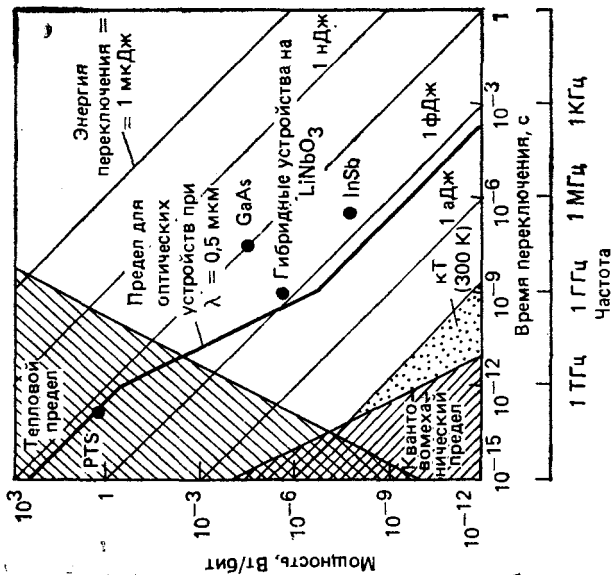
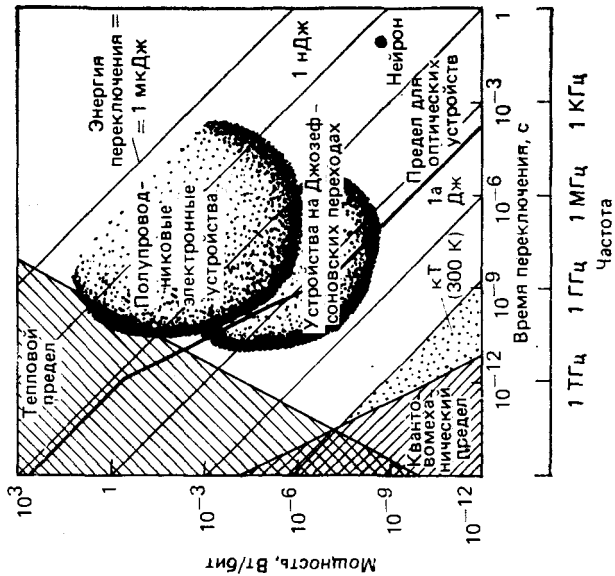


Рис. 9.1. Зависимости затрачиваемой на переключение мощности от времени переключения для различных оптических и электронных устройств [3].
PTS — полидиацилелен (нелинейный материал).

ческими. Таким образом, сохраняются преимущества и электронных, и оптических устройств. Суммарный эффект при таком подходе состоит в том, чтобы сделать возможным конструирование больших и сложных логических устройств, способных рассеивать большие мощности.

Один из способов оценить роль оптических соединений в компьютерных системах состоит в том, чтобы сравнить различные архитектуры параллельной обработки, оценивая степень сложности выполняемых задач. На рис. 9.2 изображен модифицированный вариант схемы из работы [4], иллюстрирующей потенциальные возможности оптических межэлементных соединений как функцию их числа. Из рисунка становится очевидным, что от оптики можно ожидать выполнения все более значительных задач по мере увеличения степени параллелизма обработки. Диаграмма также указывает, что степень сложности каждого обрабатываемого элемента имеет тенденцию к уменьшению по мере роста числа межэлементных соединений. В конечном итоге обрабатываемые элементы сводятся к простым вентилям, и структура обработки становится все ближе к области чисто комбинационной логики. Именно такими свойствами обладают системы, рассмотренные в данной главе. В данном случае системы не обладают памятью в традиционном смысле. И тем не менее двух- или трехуровневые комбинационные логические матрицы, позволяющие образовать логически полные наборы функций и реализовать обычную логику, могут быть классифицированы либо как устройство памяти с адресацией к месту хранения информации, либо как устройство памяти с адресацией к содержимому [5, 6]. Эти виды устройств также

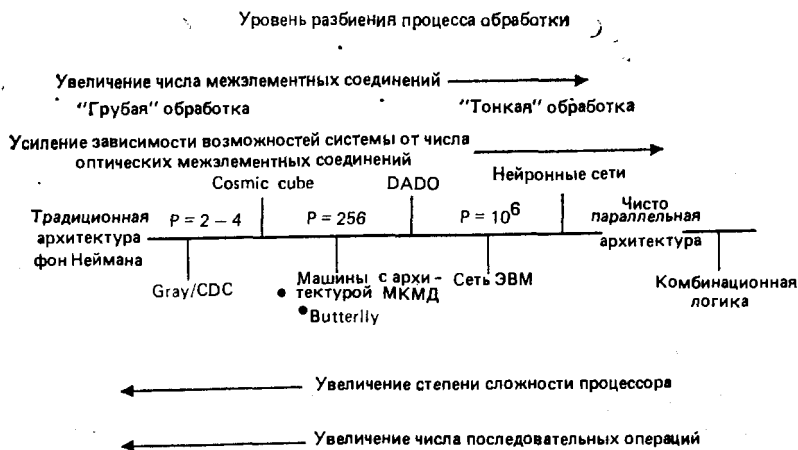


Рис. 9.2. Иллюстрация степени параллелизма систем обработки данных в зависимости от числа межэлементных соединений (взято с изменениями из [4]).

получили название памяти прямого доступа и ассоциативной памяти [7, 8]. Например, постоянное запоминающее устройство (ПЗУ) и программируемые логические матрицы (ПЛМ) соответственно являются подмножествами этих двух групп. И их динамические аналоги, а именно запоминающее устройство с произвольной выборкой (ЗУПВ) и перепрограммируемые ПЛМ, также могут быть разбиты на подобные категории. По аналогии с декодирующими устройствами ПЗУ можно рассматривать как устройство, полностью декодирующее m входных сигналов с целью получения всех возможных 2^m минтермов (элементарных конъюнктивных форм), в то время как в основе ПЛМ обычно лежат простые 1- или 2-разрядные декодеры, позволяющие получить ограниченное число элементарных конъюнктивных форм. В некотором смысле различие между ПЗУ и ПЛМ можно рассматривать как различие между устройством, способным полностью декодировать переменные входного сигнала, и устройством, способным только частично декодировать входные переменные. При описании логических функций, реализуемых ПЛМ, обычно используется перекрестная сеть соединений, задающая определенные переключения в схеме [6]. Ее применение отнюдь не является случайным. Перекрестная сеть является основным вариантом в том плане, что она обеспечивает наиболее эффективные из всех возможных одноступенчатые соединения между любыми (или всеми) N входными сигналами с любыми (или всеми) N выходными сигналами [6, 9]. Следует заметить, что каскадирование двух пространственных модуляторов света (ПМС) для получения матрицы параллельных вентилях И не обеспечивает реализации всех возможных соединений, характерных для перекрестной сети. В результате большинство логических систем, не использующих перекрестных соединений, являются весьма ограниченными в своей общности.

Со времени создания в 1978 г. Станфордском университете оптического умножителя матрицы на вектор [10] оптические перекрестные сети играли центральную роль в развитии различных аналоговых и псевдоцифровых архитектур оптических компьютеров [6, 9]. Появились перспективные оптические устройства, представляющие интерес благодаря возможностям получения высоких коэффициентов разветвления по выходу и объединения по входу, а также достижения очень высоких скоростей передачи данных [9, 11]. Недавно была разработана чисто цифровая архитектура оптоэлектронной программируемой логической матрицы (ОПЛМ), основывающаяся на перекрестной сети [6, 11—13]. Различия между аналоговыми, псевдоаналоговыми и чисто цифровыми подходами обсуждаются в [6]. Вкратце можно заметить, что чисто цифровые устройства демонстрируют заметные преимущества над аналоговыми

системами связи. Особый интерес для логического устройства представляет величина мощности, требуемая для достижения определенной производительности вычислений¹⁾. Эта величина в цифровых системах по сравнению с аналоговыми намного ниже из-за снижения требований к соотношению сигнал — шум. В настоящее время в аналоговых оптических вычислительных системах именно доступный уровень мощности в конечном счете ограничивает параллелизм. В цифровых оптических вычислительных системах с перекрестной схемой имеет место та же ситуация, за тем исключением, что в качестве критерия параллелизма можно рассматривать произведение коэффициентов объединения по входу и разветвления по выходу. Это произведение эквивалентно полному числу межэлементных соединений. Вопросы мощности, производительности и параллелизма рассмотрены более детально в разд. 9.2, где производительность обсуждается с позиции чисто комбинационной логической системы. В этом разделе также указано, что параллелизм тесно связан с полосой пропускания системы, которая снова связана с производительностью системы. Хотя производительность имеет важное значение, способность решать конкретные проблемы в конечном счете связана со сложностью вычислительной задачи. Различие между производительностью и сложностью связано с приемлемым соотношением между коэффициентами объединения по входу и разветвления по выходу. Эти различия подробно обсуждаются в разд. 9.3, где рассматривается влияние декодеров высокого порядка на сложность ПЛМ.

Были сконструированы разнообразные волоконно-оптические матрицы, основанные на сети перекрестных соединений [6, 11—13]. Эти устройства являются чисто параллельными и выполняют каждую команду за один тактовый цикл. На протяжении данной главы волоконные матрицы будут называться ОПЛМ. Их основная архитектура изображена на рис. 9.3. Традиционная ПЛМ основывается на декодере, за которым следует матрица элементов ИЛИ-И, служащая для выработки определенной логической функции. В предлагаемом подходе часть схемы с элементом ИЛИ заменяется на последовательность элементов ИЛИ-НЕ, чтобы в максимальной степени воспользоваться преимуществами оптических соединений при реализации объединения по входу и разветвления по выходу [6]. В системах этого вида коэффициент объединения по входу определяет число выходных каналов декодера, служащих входами в ПЛМ, в то время как коэффициент разветвления по выходу определяет число минимизированных термов произведения. Данный подход позволяет обойти ограничения, присущие ранним вариантам ПЛМ, построенных по схемам со свободным размещением

¹⁾ Называемая далее производительностью. — Прим. перев.

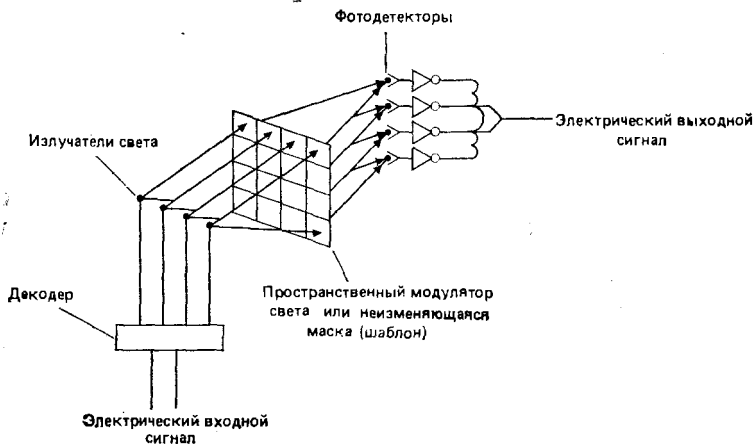
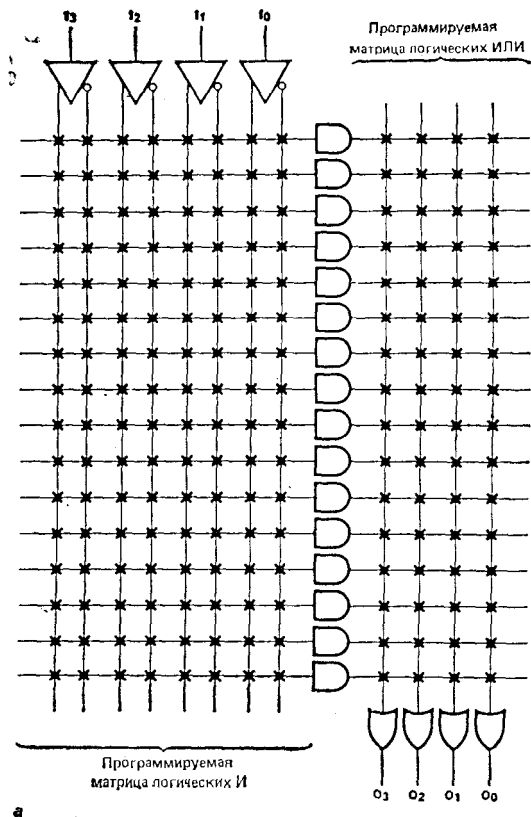


Рис. 9.3. а — схема двухуровневой ПЛМ, основанная на операциях И-ИЛИ. б — эквивалентная оптоэлектронная схема, основанная на операциях НЕ-ИЛИ-НЕ-ИЛИ.

элементов в пространстве [14] или созданных на основе интегрально-оптических схем [15], с традиционными последовательностями элементов ИЛИ-И. Хотя описанный на рис. 9.3 оригинальный вариант схемы содержит как электрические, так и оптические межэлементные соединения, ожидается, что будущие варианты будут принимать исключительно оптические входные сигналы и производить также оптические выходные сигналы [16].

Волоконные линии в общем являются более привлекательными, чем подходы, использующие голографические межэлементные соединения или соединения за счет распространения сигналов в свободном пространстве, поскольку они обеспечивают большие значения коэффициентов объединения по входу и разветвления по выходу, а также работу без перекрестных помех. Принципы организации волоконно-оптических систем с успехом могут быть использованы при существенном увеличении их размеров¹⁾. А кроме того, волоконно-оптические компоненты быстро совершенствуются вследствие быстрого распространения волоконной оптики в различных прикладных задачах техники связи. В работе [11] были рассмотрены преимущества соединений на волоконных системах над голографическими системами межэлементных соединений. В целом волоконные системы представляются весьма многообещающими, однако прежде чем делать окончательный вывод, следует рассмотреть вопрос о перестраиваемости системы с жестким «монтажом». Далее, обойти проблеме постоянных соединений нам поможет способность оптических волокон обеспечить высокие коэффициенты объединения по входу и разветвления по выходу. В разд. 9.4 будет показано, что если эти коэффициенты достаточно высоки, то может быть реализован эквивалент перестраиваемой системы [17].

Схема, показанная на рис. 9.3, была реализована на основе волоконно-оптических соединяющих элементов и приспособлена для работы в режиме МКОД — со многими потоками команд и одним потоком данных [18], хотя основная архитектура сама по себе является достаточно гибкой, чтобы работать в различных структурах. В разд. 9.2 будет показано, что независимо от структуры производительность определяется произведением коэффициентов объединения и разветвления и ширины полосы частот системы, или, что эквивалентно, произведением числа межэлементных соединений и ширины полосы частот системы. Наиболее важными факторами, ограничивающими производительность, являются рассеиваемая мощность и плотность упаковки межэлементных соединений. В самом деле, одно из прин-

¹⁾ Данный принцип далее будем называть масштабированием. — *Прим. перев.*

ципиальных преимуществ ОПЛМ состоит в способности распределять сравнительно высокие мощности по сравнительно большим площадям.

Перед тем как начать изучение основной части этой главы, полезно поразмыслить над ролью теневого метода в развитии волоконно-оптических программируемых логических матриц. Теневой метод, впервые предложенный в [19], дал возможность реализовать режим работы ОКМД, т. е. с одним потоком команд и многими потоками данных, с помощью комбинаций источников света и масок. В первоначальном варианте возможные команды были ограничены шестнадцатью возможными функциями двух двоичных входных переменных и вводились через параллельные «окна» в большой массив данных. Система тогда функционировала как совокупность большого числа очень маленьких параллельных ПЗУ или ЗУПВ. Дальнейшая разработка этой методики была выполнена этим же автором и его коллегами [8, 20] для случая более сложных процедур, основанных на большом числе двоичных входных переменных, и где аналогично описанному выше случаю, система работала как группа параллельных ПЗУ или ЗУПВ, имеющих более сложную структуру или больший масштаб выполняемых операций. Кроме того, эта группа исследователей использовала эту же систему в режиме адресации к содержанию памяти [5, 8, 21], чтобы реализовать логически минимизированные таблицы истинности, подобно тому, как это делают для ПЛМ. Используя декодеры высокого порядка для преобразования многозначных сигналов в двоичные, авторы установили, что для многозначного случая потребовалась бы значительно меньшая энергия, приходящаяся на одно переключение. Фактически, если бы они использовали двоичные декодеры высокого порядка с двоичными входными сигналами, они получили бы близкие результаты. Как уже обсуждалось ранее, взаимосвязь декодеров высокого порядка со сложностью вычислений исследуется в разд. 9.3. Хотя теневой метод и представляет интерес, но он с трудом может быть использован совместно с волокнами. Неспособность теневых систем работать в волноводном режиме жестко ограничивает практическую пригодность этой методики, тем не менее работа закладывает основу метода получения несокращенной и сокращенной таблиц истинности для различных масок. Это в конечном счете приводит к заключению, что набор масок в перекрестной схеме, в самом деле, обеспечивает наиболее эффективный из возможных способов отображения информации для программируемых логических матриц. В разд. 9.3 будут анализироваться некоторые из характеристик волоконно-оптических ОПЛМ, и, в частности, в разд. 9.2.4 будет показано, что для большого числа входных переменных использование сокращенных таблиц истинности, реализованных на ПЛМ, является

более предпочтительным, чем применение полностью декодированных таблиц истинности, реализованных с помощью ПЗУ или ЗУПВ.

9.2. Характеристики волоконно-оптических ПЛМ

9.2.1. Коэффициенты объединения по входу и разветвления по выходу

Произведение мощности на ширину полосы пропускания является важным параметром, используемым при разработке интегральных схем со сверхвысокой степенью интеграции. Это произведение также позволяет проводить сравнение оптоэлектронных логических матриц со всеми существующими электронными логическими матрицами. В последних подразделах данного раздела будет показано, что для фиксированных чувствительности фотодетектора, полосы частот и частоты появления ошибок произведение коэффициентов объединения по входу и разветвления по выходу для волоконно-оптической логической матрицы связано с мощностью входного сигнала. В дополнение к этому будет показано, что произведение этих коэффициентов оказывается связано с общим числом межэлементных соединений и производительностью системы. По этим причинам коэффициенты разветвления и объединения имеют критические значения. На рис. 9.4 показан пример соединения волокон встык, что позволяет реализовать высокие коэффициенты объединения по входу и разветвления по выходу либо в одном каскаде, либо в древовидной структуре. Данная методика была специально разработана для того, чтобы сделать возможной реализацию больших волоконно-оптических логических матриц [12]. В случае необходимости разветвления волокна одиночное волокно большего диаметра служит источником, освещающим жгут волокон, имеющих маленький диаметр. Таким образом, свет от толстого волокна распределяется по всем тонким волокнам. Исходя из предположения о том, что величины угловых апертур тол-

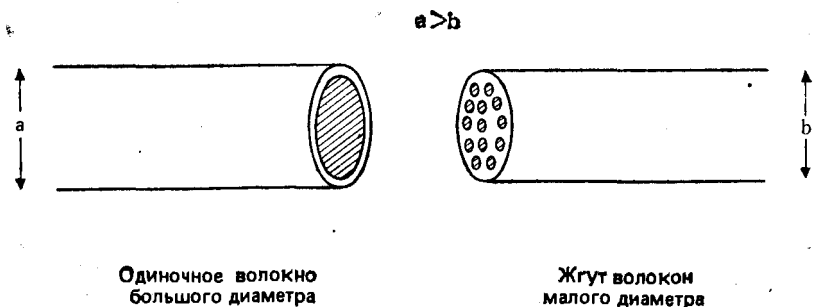


Рис. 9.4. Схема соединения «встык» волокна со жгутом волокон,

стого и тонких волокон являются постоянными, согласно теоретическим оценкам, получим, что максимальная мощность, передаваемая в каждое из тонких волокон, уменьшается пропорционально величине, на которую «не совпадают» площади торцов тонких волокон и толстого волокна. Практически передаваемая в тонкое волокно мощность оказывается несколько ниже ожидаемой вследствие потерь при соединении и неоднородного распределения света в толстом волокне. Для гауссова распределения света в волокне эффективный диаметр связи обычно ограничен половиной диаметра толстого волокна.

Из рис. 9.4 следует, что свет, распространяющийся по каждому из тонких волокон, суммируется в толстом волокне. Не учитывая потери на состыковку волокон, можно считать, что вся мощность, переносимая каждым из тонких волокон, передается в толстое волокно. Если предположить, что максимальный коэффициент объединения по входу соответствует наилучшему из всех возможных случаев, когда световую мощность передает лишь одно из тонких волокон, тогда минимальная регистрируемая мощность будет в конечном счете определяться переданной по тонкому волокну мощностью и степенью несовпадения площадей детектора и толстого волокна. При этом, естественно, считается, что угловые апертуры постоянны, а уровень регистрируемой мощности соответствует конкретному фотодетектору, определенным значениям ширины полосы пропускания и частоте появления ошибок. Это находится в хорошем соответствии с представленной в [22] теоремой, согласно которой произведение квадрата величины входной апертуры и площади входного «пятна» света не может превышать произведение величины выходной апертуры и площади выходного пятна. Предполагая, что затухание в волокне и потери при состыковке волокна не приводят к значительным потерям мощности, можно выделить четыре наиболее важных параметра, определяющие коэффициенты объединения по входу и разветвления по выходу. Такими параметрами являются мощность входного оптического сигнала, чувствительность фотодетектора, ширина полосы пропускания и частота появления ошибок. Фактически все эти параметры являются взаимосвязанными. В следующих двух подразделах будет анализироваться взаимосвязь этих параметров, а также будут рассмотрены предельные возможности, определяемые существующими ограничениями на плотность упаковки волокон и рассеиваемую мощность.

9.2.2. Рассеиваемая мощность

Для того чтобы установить достижимые на практике предельные величины коэффициентов объединения по входу и разветвления по выходу, необходимо проанализировать чувствитель-

ность фотодетектора, ширину полосы пропускания, частоту появления ошибок и передаваемые мощности. Один из путей установления предельных значений состоит в определении частоты появления ошибок, ширины полосы пропускания и коэффициентов объединения по входу и разветвления по выходу для интересующей системы. После этого величина требуемой мощности должна быть сопоставлена с типичными предельными величинами требуемой мощности для близких систем (чтобы убедиться в том, что оценки коэффициентов объединения/разветвления являются корректными). В дополнение к этому необходимая плотность упаковки волокна должна согласовываться с выбранными размерами системы.

Частота появления ошибок, по-видимому, является наиболее легко определяемым параметром, поскольку стандартом для волоконно-оптических систем является одна ошибка на миллиард битов. Чтобы достичь этой частоты появления ошибок при скорости передачи данных в один Гбит/с при условии использования высококачественных лавинных фотодиодов, требуются минимальные мощности сигналов (60 нВт). При частоте появления ошибок в 1 Гбит/с этот уровень мощности дает в среднем 300 фотонов на бит (в предположении, что число битов во включенном состоянии равно числу битов в выключенном состоянии). Если произведение коэффициентов объединения по входу и разветвления по выходу составляет 100 миллионов (каждый из коэффициентов составляет около 10 000), то требуется средняя мощность излучателя, равная 6 Вт. В соответствии с указанной выше теоремой снижение необходимой мощности может быть получено при выборе диаметра тонкого волокна менее диаметра активной области фотодетектора. Для волокна с диаметром 75 мкм типичное отношение площадей волокна и фотодетектора может составлять $1/4$, так что принципиально можно достичь снижения средней мощности излучателя до 1 Вт. На практике потери за счет состыковки волокна и неоднородности распределения световой мощности могут потребовать использования несколько больших мощностей излучателя, но влияние этих факторов может быть уменьшено путем соответствующего увеличения величины апертуры передачи света от излучателя до фотодетектора. Так как мощность излучателя в 1 Вт представляет собой практический предел для приемлемых видов излучателей, то теоретически максимальное значение произведения коэффициентов объединения по входу и разветвления по выходу составляет 100 000 000. С точки зрения возможных конструкций ОПЛМ теоретически возможно использование максимум 10 000 излучателей, 10 000 фотодетекторов и 100 000 000 межэлементных соединений.

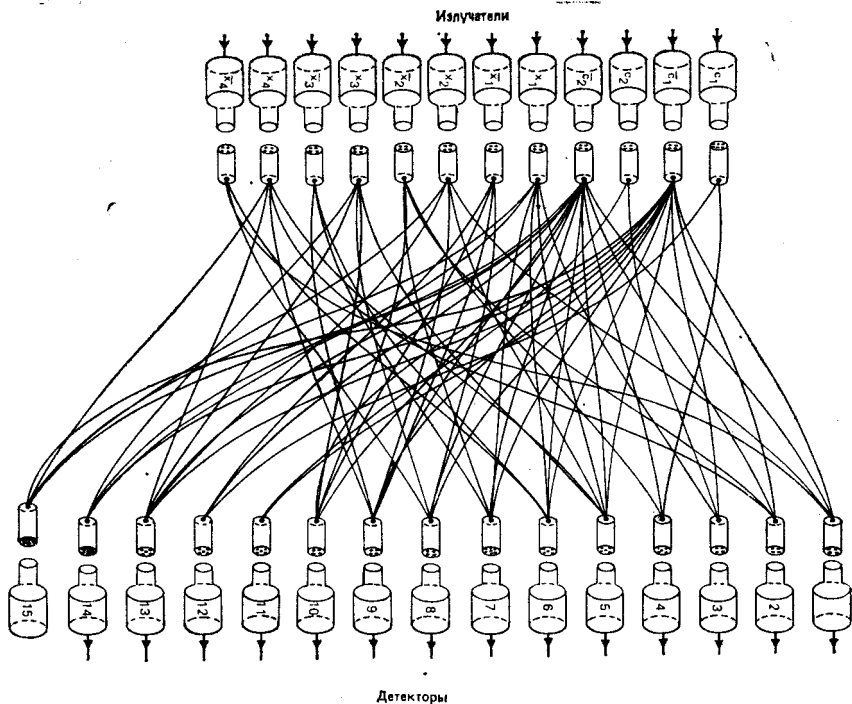
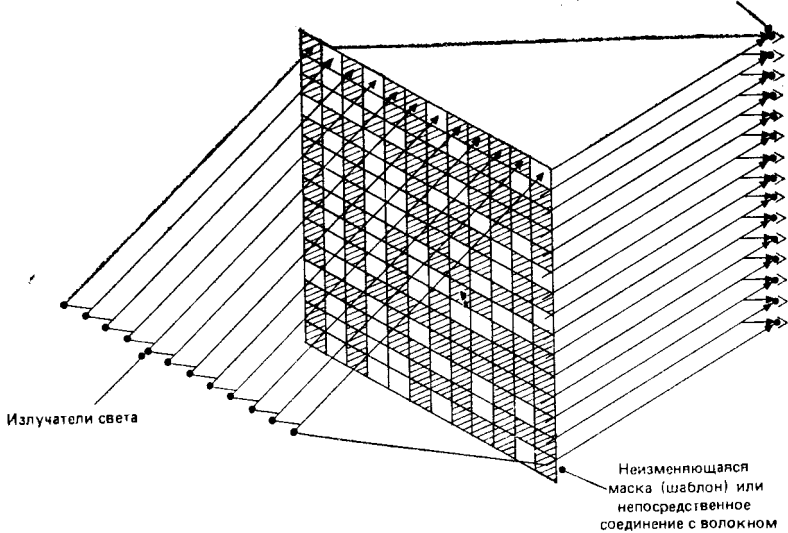
Возможно ли достижение указанных параметров на практике, можно оценить на примере модельного устройства, в кото-

ром мощность в 10 кВт рассеивалась бы на поверхности сферы, имеющей такой диаметр, что дисперсия света в волокне существенно не снижала бы ранее установленную скорость передачи данных в 1 Гбит/с. В случае передачи сигналов по многомодовому волокну с частотой 1 Гбит/с межмодовая дисперсия ограничивает линейный размер системы величиной в 9 м. Для сферы диаметром в 9 м площадь поверхности составляет 250 м². При этом требуемая скорость отвода тепла равна 40 Вт/м², что не трудно получить на практике. В самом деле, достаточно разумно предположить, что при соответствующей упаковке волокна будет возможно уменьшить диаметр сферы до величины приблизительно в 1 м. Ограничения, связанные с плотностью упаковки волокна, запрещают дальнейшее сжатие устройства. В данной ситуации требуется отвод тепла около 0,3 Вт/см², что составляет величину, вполне соответствующую возможностям обычных охлаждающих систем. Конечно, возникает и дополнительный нагрев устройства, связанный с неэффективностью преобразования электрических сигналов в световые в излучателях и фотодетекторах. В принципе этот дополнительный нагрев может быть устранен путем увеличения теплоотвода за счет использования специальных теплоотводящих жидкостей.

9.2.3. Производительность и возможность изменения масштабов устройства

Одним из наиболее значительных преимуществ волоконно-оптических матриц является их способность распределять большие мощности по сравнительно большим площадям. Рассмотрение вопроса о том, насколько может изменяться масштаб этих систем, дает возможность понять предельные возможности описанного выше подхода. В этих целях целесообразно ввести две новые характеристики, имеющие принципиальное значение для разработки программируемых логических матриц. Первая из них — это число возможных межэлементных соединений для перекрестной (или близкой к этому) сети с определенными параметрами, в то время как вторая из них — производительность самой логической матрицы. В чисто комбинационной логической системе взаимосвязь между двумя этими величинами является вполне ясной, поскольку производительность является величиной, пропорциональной произведению ширины полосы пропускания системы и числа межэлементных соединений. Ниже это отношение будет обсуждаться более детально.

На рис. 9.5, а изображена соответствующая схема перекрестного пропускания света системой (или схема маскирования), показанная для нескольких 2-разрядных команд, состоящих из операций сложения, сдвига и вычисления дополнения. Сокращенные таблицы истинности, связанные с этими командами,



Выходные сигналы, подаваемые на вентили ИЛИ

Рис. 9.5. а — схема перекрестного пропускания света для набора 2-разрядных команд [10]. б — эквивалентная волоконно-оптическая схема,

описаны, например, в [12]. Эквивалентный случай непосредственного соединения волокон показан на рис. 9.5, б. Из данного рисунка следует, что число двоичных битов, необходимое для передачи информации о заданных командах, пропорционально числу элементов в маске. Несомненно, что максимально возможное число межэлементных соединений также является пропорциональным числу этих элементов. Поскольку произведение коэффициентов разветвления по выходу и объединения по входу определяет число элементов в устройстве, то произведение этих коэффициентов будет пропорциональным максимально возможному числу соединений. На протяжении данной главы под числом межэлементных соединений будет подразумеваться максимально возможное число соединений в перекрестной сети. Так, перекрестная сеть из $N \times N$ элементов обеспечивает возможность создания N^2 межэлементных соединений. Рис. 9.5, а служит иллюстрацией к тому факту, что перекрестная сеть обладает присущим ей параллелизмом, связанным с процессами объединения по входу и разветвления по выходу. Так как коэффициенты разветвления по выходу и объединения по входу связаны с четырьмя определенными выше параметрами, то в этом же контексте должен быть рассмотрен и вопрос о параллелизме рассматриваемых систем. Аналогичная ситуация наблюдается в аналоговых системах, где степень параллелизма выполняемых операций тесно связана с мощностью входного сигнала, шириной полосы частот, чувствительностью детектора и необходимым значением отношения сигнал/шум.

Для того чтобы выяснить возможность изменения масштабов устройства, полезно положить постоянной рассеиваемую в системе мощность. Это может быть сделано в предположении, что мощность, выделяемая на единицу площади поверхности, постоянна. В результате мощность зависит от длины как L^2 , а произведение мощности на ширину полосы частот будет пропорционально L в предположении, что именно ширина полосы частот волокна (обратно пропорциональная L) ограничивает характеристики системы. Другим важным параметром, описывающим возможность масштабирования системы, является относительная чувствительность детектора, обратно пропорциональная полосе частот. Чувствительность детектора тогда масштабируется пропорционально L , так что произведение мощности и чувствительности детектора оказывается пропорционально L^3 . Из ранее рассмотренных в этом разделе свойств коэффициентов объединения по входу и разветвления по выходу становится ясно, что произведение мощности и чувствительности детектора пропорционально произведению коэффициента объединения по входу и разветвления по выходу. По существу, если почему-либо возрастают коэффициенты разветвления по выходу источника мощности и объединения по входу детектора, то для по-

лучения заданной частоты появления ошибок в детекторе требуется большая чувствительность детектора. Для того чтобы произведение коэффициентов разветвления по выходу и объединения по входу было пропорционально L^3 , диаметр волокна должен быть достаточно мал, чтобы волокна не соприкасались между собой; в противном случае при изменении масштаба системы произведение коэффициентов разветвления по выходу и объединения по входу будет определяться зависимостью L^2 . Данный критерий недопустимости соприкосновения волокон называется пределом одной моды и будет подробно обсуждаться в этом же разделе.

На основе приведенных выше данных соотношения масштабирования для производительности чисто комбинационной логической системы могут быть определены сравнительно просто. Ранее в этом разделе было указано, что конфигурация элементов изображения, приведенного на рис. 9.5, непосредственно получается с помощью сокращенной таблицы истинности. Здесь число строк, или коэффициент разветвления по выходу, определяет минимизированное число изображений, создаваемых ПЛМ, или число термов произведения (логического), в то время как число столбцов, или коэффициент объединения по входу, определяет число выходных каналов декодера, служащих входными каналами ПЛМ. Из сказанного выше очевидно, что число элементов изображения, необходимых для реализации отображения исходного изображения, определяет физическую «емкость» соответствующей ПЛМ. В таком случае произведение числа элементов N и ширины полосы частот B дает критерий для измерений производительности системы. Если для конкретной операции или при расчетах, выполняемых с помощью логической матрицы, потребуется большее число тактовых циклов C или меньшее число ячеек P , то из отношения NB/PC получим величину пропускной способности системы, измеряемой числом операций в секунду (как сообщалось, например, в [7, 8]). В ином варианте производительность системы может быть задана либо как произведение коэффициентов разветвления по выходу и объединения по входу и ширины полосы частот, либо как произведение числа межэлементных соединений на ширину полосы частот. Третий и эквивалентный способ оценки производительности заключается в анализе коэффициента, получаемого при перемножении мощности, чувствительности детектора и ширины полосы частот. Все три подхода указывают, что производительность масштабируется пропорционально.

На рис. 9.6, *a* представлена зависимость числа соединений от ширины полосы пропускания, в то время как на рис. 9.6, *b* показана зависимость относительной производительности от линейного размера системы. Основываясь на аргументах, указан-

ных ранее в этом разделе в связи с условиями рассеяния тепла, можно получить, что 10^8 межэлементных соединений, работающих со скоростью 1 Гбит/с, могут быть размещены в сфере диаметром 1 м, что дает относительную производительность 10^{17} . Эти данные точками показаны на рис. 9.6, а, б, в то время как остальные значения получены из ранее высказанных соображений по поводу возможности масштабирования системы. По мере увеличения размера волоконной системы до 100 м параметры ее быстродействия приближаются к пределу одномодового волокна, когда дальнейшее уменьшение диаметра волокна невозможно. Эта величина условно принята за 10 мкм. Здесь число соединений ограничено плотностью упаковки и производительность масштабируется пропорционально L , а не L^2 , в то время как число межэлементных соединений масштабируется пропорционально L^2 , а не L^3 . В следующей части этого раздела будет показано, что относительная производительность лучших из ныне существующих электронных логических матриц с плот-

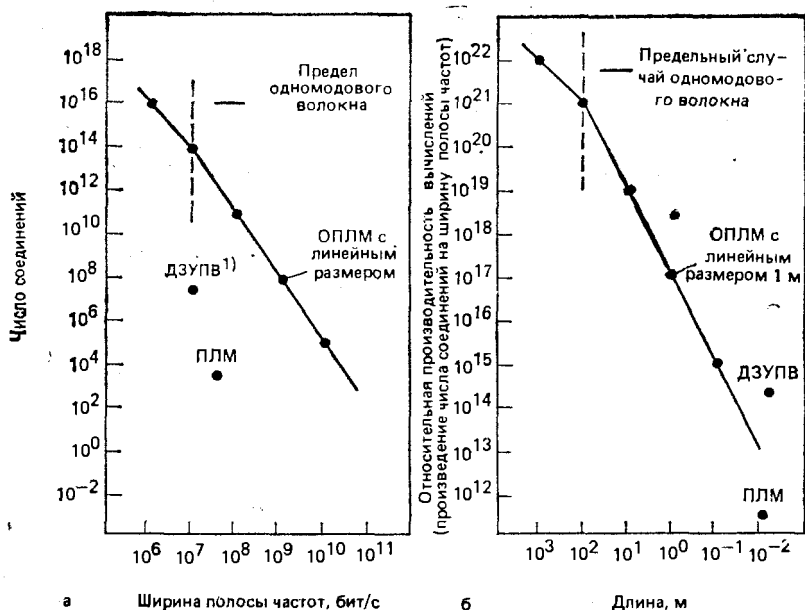


Рис. 9.6. а — зависимость максимального числа межэлементных соединений от ширины полосы частот при параметрах рассеяния мощности и плотности упаковки, определяемых возможностями волоконно-оптических ПЛМ.

¹⁾ ДЗУПВ — динамическое запоминающее устройство с произвольной выборкой.

б — зависимость относительной производительности вычислений от общей длины устройства (для тех же случаев, что приведены на рис. 9.6, а).

ной упаковкой составляет величину порядка 10^{14} , что на семь порядков меньше, чем дает оценка, выполненная в пределе одномодового волокна.

9.2.4. Сравнение возможностей различных методик реализации комбинационной логики

Ранее в данном разделе было показано, что коэффициенты объединения по входу и разветвления по выходу взаимосвязаны с шириной полосы пропускания системы. Одно из преимуществ применения ОПЛМ заключается в том, что использование отдельных электронных компонент с волоконно-оптическими межэлементными соединениями позволяет использовать очень низкие плотности размещения реальных вентилях по сравнению с интегральными схемами со сверхвысокой степенью интеграции. Таким образом, в значительных по объему волоконно-оптических системах может быть рассеяна существенно большая мощность, чем в компактных интегральных схемах со сверхвысокой степенью интеграции. Так как скорости переключения в конечном счете ограничены возможностью отвода выделяемого тепла, волоконно-оптические системы позволяют, кроме того, достичь более высоких тактовых частот. В самом деле, обычным явлением стала работа волоконно-оптических систем в области гигагерцевых частот, тогда как очень трудно добиться работы микроэлектронного чипа со сверхвысокой степенью интеграции при таких частотах.

На рис. 9.7, а проводится сравнение ожидаемых предельных характеристик ОПЛМ, имеющих общую длину устройства в 1 м, а также параметров ряда промышленно освоенных устройств: динамических запоминающих устройств с произвольной выборкой (ДЗУПВ), статических запоминающих устройств с произвольной выборкой (СЗУПВ), матричных устройств программируемой логики (ПЛМ), логических устройств, программируемых пользователем в зависимости от условий эксплуатации (ЛУПУЭ). Сравнение основано на зависимости скорости переключения от числа переменных во входном сигнале. На рис. 9.7, б представлены сравнительные данные, характеризующие зависимости скорости переключения от числа термов произведения. В данном сравнении не делается различий между устройствами, основанными на использовании логически минимизированных термов произведения и соответственно не минимизированных термов произведения. Наибольшей плотностью упаковки элементов обладают ДЗУПВ. В самом деле, эти устройства можно рассматривать как пример изделий, находящихся на переднем крае современной технологии изготовления чипов. Хотя ДЗУПВ обычно уже обладают внутренним разделением каналов, может быть предложена система с числом раз-

личных переключаемых каналов, равным произведению коэффициентов разветвления по выходу и объединения по входу (разработанная ранее для систем с перекрестной схемой). Математически это эквивалентно произведению числа переменных входного сигнала на число булевых термов произведения. Для ДЗУПВ емкостью 1 Мбит, имеющего 20 входных каналов, 10^6 термов произведения и работающего с частотой 10 МГц (20 Мбит/с), относительная производительность составляет $4 \cdot 10^{14}$. Следует заметить, что эта величина относится к крем-

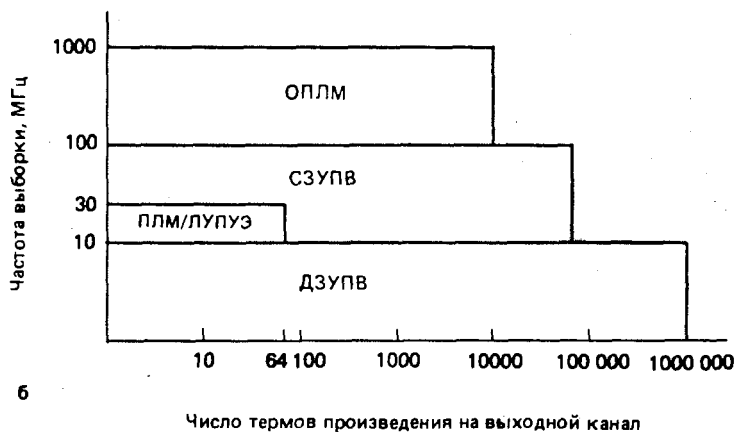
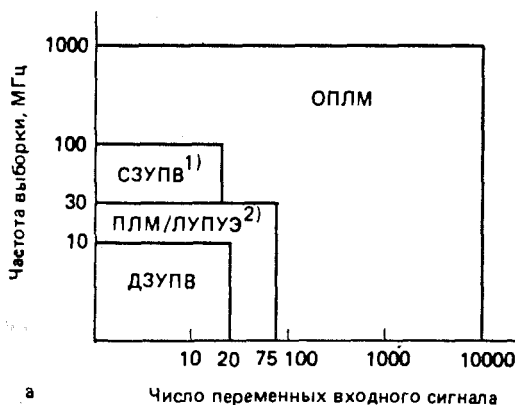


Рис. 9.7. а — зависимость максимальной скорости выборки от числа переменных входного сигнала для различных типов логических матриц. Сравнение производится для линейного размера ОПЛМ в 1 м.

¹⁾ СЗУПВ — статическое запоминающее устройство с произвольной выборкой.

²⁾ ЛУПУЭ — логическое устройство, программируемое в зависимости от условий эксплуатации.

б — зависимость максимальной скорости выборки от числа термов произведения для тех же логических матриц. Не учитывается различие между обычными и минимизированными произведениями.

ниевым приборам. Из-за сложностей с теплоотводом логические матрицы, изготовленные из арсенида галлия, обычно имеют по крайней мере на два порядка меньшую производительность, хотя потенциально их быстродействие на один или два порядка выше, чем для кремниевых приборов. В результате суммарный выигрыш в производительности, если и существует, то оказывается очень незначительным. Для сравнения заметим, что в одномодовом пределе ОПЛМ теоретически способны обеспечить относительную производительность, равную 10^{21} .

Из рис. 9.7, а следует, что при числе переменных входного сигнала более двадцати более целесообразным становится использование устройств программируемой матричной логики, программируемых пользователем логических устройств, программируемых логических матриц, а применение ДЗУПВ становится менее выгодным. Даже для лучшего из существующих ДЗУПВ, имеющего 75 входных каналов и 64 термина производства, работающего при 30 МГц, относительная производительность составляет только $1,4 \cdot 10^{11}$. И скорее всего (основываясь на этой характеристике) волоконно-оптические ОПЛМ ввиду их явного преимущества начнут вытеснять с рынка электронные изделия.

Весьма спорным является вопрос о возможности создания специальных электронных систем, использующих такие оконечные передающие линии, которые позволили бы достичь значения коэффициента разветвления по выходу, сравнимого с возможностями ОПЛМ (для той же полосы частот). Анализ зависимостей плотности упаковки от скорости выборки и коэффициента разветвления по выходу от скорости выборки [23] указывает на то, что стоимость и размеры таких систем были бы неприемлемыми; более того, при обсуждении ОПЛМ даже и не затрагивался вопрос о потенциальной способности этих систем реализовать одновременно высокие коэффициенты объединения по входу и разветвления по выходу. В самом деле, в большинстве современных сетей связи с самыми различными степенями сложности уже отказались от концепций применения передающих линий в пользу волоконно-оптических систем.

9.3. Проблемы создания логических схем для волоконно-оптических систем

9.3.1. Методы минимизации и оптимального присваивания переменных

История развития методов логической минимизации является и длинной, и достаточно интересной. Процесс минимизации с точки зрения процедуры вычислений является настолько сложным, что, даже используя для выполнения этой задачи супер-

компьютер, все равно требуется вводить более 30 переменных. Процесс поиска все более эффективных подходов начался разработкой таких точных методов, как алгоритмы Квина—Мак-Класки [24] и Шейнманна [25], и привел к созданию таких эвристических подходов, как MINI [26] и ESPRESSO [27]. В работе [28] было установлено, что для определенных видов функций использование 2-разрядных декодеров с оптимальным присваиванием переменных может быть гораздо более эффективным, чем применение традиционных двухуровневых ПЛМ, т. е. ПЛМ с 1-разрядным декодером. В качестве примера следует заметить, что для операции n -разрядного сложения без переноса цифр в разрядах и 1-разрядного декодера число минимизированных термов экспоненциально зависит от n , тогда как для 2-разрядного декодера и при оптимальном присваивании переменных число термов пропорционально n^2+1 . Для 2-разрядного декодера оптимальное присвоение переменных представляет собой просто процедуру правильного попарного «соединения» входных сигналов в каждом декодере, т. е. попарное сложение соответствующих битов от двух складываемых чисел, а не просто сложение произвольных 2 битов в пределах каждого из чисел. Кроме того, для еще большего уменьшения числа логических элементов в системе фазы выходных сигналов часто тоже могут быть оптимизированы. Для случая применения декодеров, позволяющих использовать m многозначных входных сигналов и обеспечивающих формирование двоичных выходных сигналов (т. е. декодеров, работающих с основанием p и позволяющих преобразовать p в p^m), проблема оптимального присвоения переменных становится одной из основных задач кодирования входного сигнала [29].

В противоположность случаю ПЛМ с 1-разрядным декодером ПЗУ полностью декодирует m входных сигналов, создавая все возможные 2^m минтермов (элементарных конъюнктивных форм). Этот тип устройств фактически требует использования только одной комбинации логических элементов. С точки зрения математики различие между ПЗУ и ПЛМ может быть рассмотрено как разница между устройствами, способными проводить полное декодирование входных переменных, и устройствами, способными проводить только частичное декодирование переменных входного сигнала. Цель, которую преследовали авторы при написании данного раздела, заключается в исследовании промежуточной области между ПЗУ и обычными ПЛМ путем анализа влияния возможности декодеров высоких порядков на число комбинаций логических элементов в ПЛМ. В следующих двух частях этого раздела будет показано, что, хотя число комбинаций, необходимых для реализации конкретной функции, всегда монотонно уменьшается с увеличением порядка декодера, необходимые производительность вычислений

и входная мощность фактически проходят через абсолютный минимум. Для успешной разработки эффективных устройств, таким образом, требуется, чтобы степень сложности декодера была оптимизирована на основе анализа всех возможных комбинаций.

9.3.2. Роль свойств декодеров высоких порядков

Полная таблица истинности для двоичной системы с m входами содержит 2^m строк, по одной на каждую возможную комбинацию входных сигналов. Обозначая номер строки n , получим, что полное число возможных функций выходного сигнала по оценкам составляет ошеломляющую величину — 2^n . Степень сложности этих функций различается весьма значительно. Один из способов определения степени сложности функций заключается в проведении для этих функций процедуры логической минимизации и сравнения числа полученных вариантов. Это число также позволяет определить требуемый коэффициент разветвления по выходу. Термин «функциональная сложность» уместен лишь для двузначных ПЛМ, т. е. для ПЛМ с 1-разрядным декодером, и он не подходит для используемых декодеров высших порядков. Для случая декодеров высших порядков необходимо дать определение дополнительной величине, получившей название «сложности вычислений». Это понятие будет применяться для обозначения минимизированного числа логических функций, получаемых в случае использования n -разрядных декодеров. Представленные ниже данные позволят продемонстрировать тот факт, что для определенного уровня функциональной сложности сложность вычислений также может значительно различаться (в том случае, если используются декодеры высших порядков).

Чтобы оценить целесообразность применения декодеров высших порядков с точки зрения вычислительной сложности, удобно работать с набором функций, имеющих предсказуемый рост сложности по мере увеличения числа входных переменных. Одним из примеров таких наборов функций являются пороговые функции. Эти функции представляют интерес и по той причине, что они могут быть использованы для реализации обычной логики. Математический аппарат пороговой логики принципиально отличается от булевой алгебры, тем не менее теоретически возможно получить любую булеву функцию в рамках подхода пороговой логики. Данный метод является привлекательным, так как он может привести к значительной экономии числа логических вентилях и снижению требований к числу межэлементных соединений. К сожалению, недостаточный уровень развития универсальных методик получения пороговых функций ограничил степень практической полезности этого подхода [30];

тем не менее с его помощью может быть оптически реализован ряд булевых функций [31]. И наоборот, возможно получить любую пороговую функцию с помощью методов минимизации в стандартной булевой логике. Указанный процесс логической минимизации представляет собой универсальную методику синтеза, обычно называемую синтезом в исследовании операций. Недавно на основе ОПЛМ был разработан и реализован полностью программируемый генератор пороговых функций (называемый также программируемым вентилем) [13].

Пороговая функция может быть получена при вычислении внутреннего произведения весового вектора и вектора входного сигнала, а также порогового кодирования результата. Эту функцию удастся получить путем ограничения входных сигналов двоичными числами и полагая все весовые множители равными единице; в этом случае она равна сумме входных сигналов, подвергнутых пороговому кодированию. Для любого произвольного числа переменных входного сигнала данная функция может быть получена с помощью методов минимизации обычной булевой логики, что дает определенное число комбинаций или термов произведения. Одним из забавных свойств пороговых функций, как было замечено автором данной главы, является то, что один или большее число термов произведения, полученных за счет приравнивания всех весовых коэффициентов единице, представляет собой не что иное, как одну из возможных пороговых подфункций. При этом пороговые подфункции могут быть получены в предположении, что любая комбинация весовых коэффициентов принимает значения либо 0, либо 1. Тогда случай единичных весовых коэффициентов представляет максимально возможную функциональную сложность для случая 1-разрядных весовых коэффициентов. В табл. 9.1 представлен ряд значений минимизированных термов произведения для случая 1-разрядных входных сигналов с единичными весовыми коэффициентами, являющимися функциями полного числа переменных входного сигнала, значений порога и степени сложности декодера.

Данные, представленные в табл. 9.1, вполне заслуживают внимания, поскольку представляют собой первую попытку строгого исследования процесса синтеза ряда пороговых функций в рамках булевой логики, применительно к декодерам с увеличивающейся степенью сложности. Эти данные были получены в результате длительных расчетов по алгоритму ESPRESSO. При проверке данных видно, что для определенного числа входных переменных и 1-разрядных декодеров необходимое число минимизированных термов произведения, связанных с каждым из возможных значений порога, может быть получено с помощью биномиальных коэффициентов. Это соответствует числу возможных комбинаций из M переменных, обо-

Таблица 9.1. Логически минимизированные термы произведений, представленные в зависимости от степени сложности декодера, для пороговых функций 4, 8, 12 и 16 переменных входного сигнала. Точность переменных ограничена 1 бит, а соответствующие весовые коэффициенты выбраны равными 1

Число входов сигнала m	Размер декодера n	Значения пороговой функции											Кoeffициент разветвления по выводу	Относительный коэффициент объединения по входу	Относительное произведение коэффициентов разветвления по выводу и объединения по входу																																																																						
		0,5	1,5	2,5	3,5	4,5	5,5	6,5	7,5	8,5	9,5	10,5				11,5																																																																					
4	1	<table border="1"> <tr> <td>4</td> <td>6</td> <td>4</td> <td>1</td> <td colspan="8"></td> </tr> <tr> <td>2</td> <td>3</td> <td>2</td> <td>1</td> <td colspan="8"></td> </tr> <tr> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td colspan="8"></td> </tr> </table>											4	6	4	1									2	3	2	1									1	1	1	1									15	1	15																																		
4	6												4	1																																																																							
2	3												2	1																																																																							
1	1	1	1																																																																																		
4	2												8	1	8																																																																						
4	4												4	2	8																																																																						
8	1	<table border="1"> <tr> <td>8</td> <td>28</td> <td>56</td> <td>70</td> <td>56</td> <td>28</td> <td>8</td> <td>1</td> <td colspan="5"></td> </tr> <tr> <td>4</td> <td>10</td> <td>16</td> <td>19</td> <td>16</td> <td>10</td> <td>4</td> <td>1</td> <td colspan="5"></td> </tr> <tr> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> <td colspan="5"></td> </tr> <tr> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td colspan="5"></td> </tr> </table>											8	28	56	70	56	28	8	1						4	10	16	19	16	10	4	1						2	3	4	5	4	3	2	1						1	1	1	1	1	1	1	1						255	1	255																		
8	28												56	70	56	28	8	1																																																																			
4	10												16	19	16	10	4	1																																																																			
2	3												4	5	4	3	2	1																																																																			
1	1	1	1	1	1	1	1																																																																														
8	2												80	1	80																																																																						
8	4												24	2	48																																																																						
8	8												8	16	128																																																																						
12	1	<table border="1"> <tr> <td>12</td> <td>66</td> <td>220</td> <td>495</td> <td>792</td> <td>924</td> <td>792</td> <td>495</td> <td>220</td> <td>66</td> <td>12</td> <td>1</td> <td colspan="2"></td> </tr> <tr> <td>6</td> <td>21</td> <td>50</td> <td>90</td> <td>126</td> <td>141</td> <td>126</td> <td>90</td> <td>50</td> <td>21</td> <td>6</td> <td>1</td> <td colspan="2"></td> </tr> <tr> <td>3</td> <td>6</td> <td>10</td> <td>15</td> <td>18</td> <td>19</td> <td>18</td> <td>15</td> <td>10</td> <td>6</td> <td>3</td> <td>1</td> <td colspan="2"></td> </tr> <tr> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>6</td> <td>5</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> <td colspan="2"></td> </tr> <tr> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td colspan="2"></td> </tr> </table>											12	66	220	495	792	924	792	495	220	66	12	1			6	21	50	90	126	141	126	90	50	21	6	1			3	6	10	15	18	19	18	15	10	6	3	1			2	3	4	5	6	7	6	5	4	3	2	1			1	1	1	1	1	1	1	1	1	1	1	1			4095	1	4095
12	66												220	495	792	924	792	495	220	66	12	1																																																															
6	21												50	90	126	141	126	90	50	21	6	1																																																															
3	6												10	15	18	19	18	15	10	6	3	1																																																															
2	3												4	5	6	7	6	5	4	3	2	1																																																															
1	1												1	1	1	1	1	1	1	1	1	1																																																															
12	2												728	1	728																																																																						
12	4												124	2	248																																																																						
12	6												48	16/3	256																																																																						
12	12												12	512/3	2048																																																																						
16	2	<table border="1"> <tr> <td>8</td> <td>36</td> <td>112</td> <td>266</td> <td>504</td> <td>784</td> <td>1016</td> <td>1107</td> <td>1016</td> <td>784</td> <td>504</td> <td>266</td> <td>112</td> <td>36</td> <td>1</td> </tr> <tr> <td>4</td> <td>10</td> <td>20</td> <td>35</td> <td>52</td> <td>68</td> <td>80</td> <td>85</td> <td>80</td> <td>68</td> <td>52</td> <td>35</td> <td>20</td> <td>10</td> <td>1</td> </tr> <tr> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> <td>8</td> <td>7</td> <td>6</td> <td>5</td> <td>4</td> <td>3</td> <td>1</td> </tr> <tr> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> </table>											8	36	112	266	504	784	1016	1107	1016	784	504	266	112	36	1	4	10	20	35	52	68	80	85	80	68	52	35	20	10	1	2	3	4	5	6	7	8	9	8	7	6	5	4	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6560	1	6560										
8	36												112	266	504	784	1016	1107	1016	784	504	266	112	36	1																																																												
4	10												20	35	52	68	80	85	80	68	52	35	20	10	1																																																												
2	3												4	5	6	7	8	9	8	7	6	5	4	3	1																																																												
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1																																																																							
16	4												624	2	1248																																																																						
16	8												80	16	1280																																																																						
16	16												16	2048	32768																																																																						

значенному m и соответствующему некоторому моменту времени. Вычислив соответствующие булевы выражения для каждого из этих случаев, можно показать, что число необходимых термов произведения для декодеров высших порядков также связано с биномиальными коэффициентами. Возможно столь же интересным является свойство симметрии, характеризующее отдельные термы произведения. Любопытно заметить, что число термов произведения, вычисленных вдоль определенной строки значений порога, обладает осевой симметрией относительно значения в центре строки. Исследуя отдельные коэффициенты, можно показать, что обращение фазы для значений порога, меньших, чем в центре строки, дает ярко выраженную симметрию, вследствие чего можно минимизировать полное число термов произведения уже за пределами значений, представленных в табл. 9.1.

Для каждой группы переменных входного сигнала возможные пороговые значения могут рассматриваться как дополнительные входы в большую таблицу истинности, образующую макрофункцию. Эта макрофункция является в значительной мере программируемой униполярной пороговой функцией [13]. Данная макрофункция, связанная с каждой группой переменных входного сигнала, представляет определенный уровень функциональной сложности. Сложность вычислений, требуемая для синтеза данной функции, может быть определена путем суммирования всех произведений вдоль определенной строки в табл. 9.1. Результаты указаны в столбце, именуемом «коэффициент разветвления по выходу». Из представленных в данном столбце данных становится очевидным, что сложность вычислений коэффициента разветвления по выходу, связанная с каждым значением переменных входного сигнала, уменьшается монотонно с ростом сложности декодера. Как отмечалось ранее, не является удивительным тот факт, что число термов произведения должно в конечном счете равняться одному терму на один выходной канал в том случае, когда входной сигнал полностью декодируется. Один из негативных моментов, связанный с использованием декодеров высших порядков, заключается в сопутствующем увеличении коэффициента объединения по входу. В следующей части раздела будет показано, что существует оптимальный уровень сложности декодера, связанный с достижением компромисса между коэффициентами объединения по входу и разветвления по выходу. Этот оптимальный уровень сложности декодера задает минимум требований в отношении сложности вычислений, сводя к минимуму затраты мощности и энергии на проведение конкретных вычислений.

9.3.3. Производительность, функциональная сложность и затраты энергии

Для каждого значения переменных входных сигналов, показанных в табл. 9.1, связанная с ним программируемая униполярная пороговая функция представляет определенный уровень функциональной сложности. Использование декодеров высших порядков обеспечивает наличие механизма уменьшения сложности вычислений, требуемой для получения пороговых функций. В действительности это приводит к увеличению относительного коэффициента объединения по входу ПЛМ, уменьшая между тем коэффициент разветвления по выходу. Фактически именно произведение коэффициентов разветвления и объединения имеет критическое значение. При постоянной ширине полосы частот это произведение пропорционально производительности, а также мощности входного сигнала, или энергии, необходимой для полного завершения всей процедуры вычислений. Для того чтобы минимизировать соотношение между коэффициентами объединения по входу и коэффициентами разветвления по выходу, необходимо более детально рассмотреть относительный коэффициент объединения по входу ПЛМ.

Оценка относительных коэффициентов объединения по входу ПЛМ как функции сложности декодера начинается с установления того факта, что число выходных каналов для n -разрядного декодера равняется 2^n . Относительный коэффициент объединения по входу ПЛМ может быть определен путем установления числа переменных входного сигнала и последующего вычисления произведения соответствующего числа декодеров и числа выходных каналов, приходящихся на декодер. Таким образом, функция восьми переменных требует 2-разрядного декодера или двух 4-разрядных декодеров. Относительный коэффициент объединения по входу ПЛМ, следовательно, будет больше в случае 4-разрядных декодеров, чем для случая 2-разрядных декодеров. По мере увеличения сложности декодеров входного сигнала относительный коэффициент объединения по входу ПЛМ также возрастает. Единственным исключением из этого случая является переход от 1-разрядных декодеров к 2-разрядным декодерам, где относительный коэффициент объединения по входу остается постоянным.

В табл. 9.1 суммируются относительные коэффициенты объединения по входу для четырех уровней функциональной сложности, начиная с четырех переменных и кончая шестнадцатью. Относительные произведения коэффициентов разветвления по выходу и объединения по входу для каждого уровня функциональной сложности и для каждого уровня сложности декодеров вычислены и представлены в последнем столбце табл. 9.1. На рис. 9.8 представлены зависимости относительного произ-

ведения коэффициентов разветвления по выходу и объединения по входу от числа переменных входного сигнала. Из данного рисунка следует, что применение декодеров высшего порядка значительно снижает требования к относительному произведению коэффициентов разветвления и объединения, или, что эквивалентно, относительной производительности, требуемой на каждом уровне функциональной сложности. Во всех случаях требуемая производительность вычислений экспоненциально возрастает при увеличении функциональной сложности, но скорость возрастания тем не менее является значительно меньшей, чем в случае, когда используются декодеры высших порядков.

Рисунок 9.8 вводит читателя в заблуждение в том смысле, что он выполнен в предположении, что требуемая производительность для определенного уровня функциональной сложности уменьшается монотонно с увеличением сложности декодеров. Тщательный анализ последнего столбца в табл. 9.1 показывает, что относительное произведение коэффициентов раз-

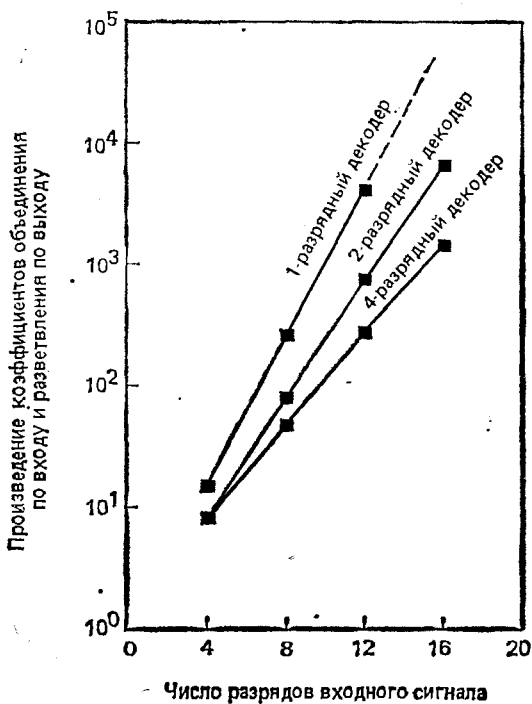


Рис. 9.8. Зависимость произведения относительных коэффициентов объединения в по входу и разветвления по выходу от числа разрядов (числа переменных) входного сигнала, приведенная для ОПЛМ с декодерами высших порядков. Указанная зависимость эквивалентна зависимости относительной производительности вычислений от степени функциональной сложности.

ветвления по выходу и объединения по входу, рассматриваемое в качестве функции сложности декодера, фактически проходит через минимум. Из этих данных становится очевидным, что существует оптимальный уровень сложности декодеров, необходимый для минимизации требуемой производительности. Следовательно, при постоянной ширине полосы частот мощность и энергия, необходимые для проведения конкретного вычисления, также могут быть минимизированы за счет соответствующего подбора сложности декодера. Фактический коэффициент на практике зависит от распределения нулей и единиц в отдельных терминах произведения. В результате этого требования к реальным коэффициентам могут быть значительно снижены по сравнению с представленными здесь номинальными пределами. Различие между номинальными и фактическими коэффициентами обсуждается в следующем подразделе.

9.4. Перестраиваемые волоконные системы

9.4.1. Волоконно-оптические перекрестные переключатели

Способность волоконно-оптических программируемых логических матриц рассеивать большие количества энергии по значительным площадям представляет собой принципиальное преимущество над электронными методами. В разд. 9.2 показано, что ОПЛМ потенциально пригодны к работе с высокими скоростями и большими значениями коэффициентов разветвления по выходу и объединения по входу. В данном разделе будет показано, что способность ОПЛМ работать в режиме с большими значениями коэффициентов создает механизм для реализации перестраиваемости систем, делая ОПЛМ прекрасными кандидатами для использования в качестве логических перекрестных переключателей. В представленных здесь разработках используется то преимущество, что достаточно большие ПЛМ могут выполнять все логические комбинации, необходимые для воздействия на эквивалентные входные-выходные установочные параметры традиционных перекрестных переключателей [11].

В идеальной параллельной системе обработки информации любой процессор должен обладать способностью одновременно связываться с любым числом других процессоров. Универсальный перекрестный переключатель представляет собой идеализированную сеть, способную устанавливать перестраиваемые соединения между любыми из N входов и любыми из N выходов без возникновения конфликтных ситуаций. Практически трудно и дорого сконструировать электронные перекрестные сети размером более 16×16 из-за ограничений, накладываемых на коэффициенты объединения по входу и разветвления по вы-

Таблица 9.2. Сокращенная таблица истинности для перекрестного переключателя 4×4

C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3	Y_4
—	—	—	—	—	—	0	0	1	—	—	—	1	0	0	0
—	—	—	—	—	—	0	1	—	1	—	—	1	0	0	0
—	—	—	—	—	—	1	0	—	—	1	—	1	0	0	0
—	—	—	—	—	—	1	1	—	—	—	1	1	0	0	0
—	—	—	—	0	0	—	—	1	—	—	—	0	1	0	0
—	—	—	—	0	1	—	—	—	1	—	—	0	1	0	0
—	—	—	—	1	0	—	—	—	—	1	—	0	1	0	0
—	—	—	—	1	1	—	—	—	—	—	1	0	1	0	0
—	—	0	0	—	—	—	—	1	—	—	—	0	0	1	0
—	—	0	1	—	—	—	—	—	1	—	—	0	0	1	0
—	—	1	0	—	—	—	—	—	—	1	—	0	0	1	0
—	—	1	1	—	—	—	—	—	—	—	1	0	0	1	0
0	0	—	—	—	—	—	—	1	—	—	—	0	0	0	1
0	1	—	—	—	—	—	—	—	1	—	—	0	0	0	1
1	0	—	—	—	—	—	—	—	—	1	—	0	0	0	1
1	1	—	—	—	—	—	—	—	—	—	1	0	0	0	1

ходу; тем не менее были созданы большие электронные перекрестные переключатели размером 256×256 [9]. Хотя для электронных перекрестных переключателей удается достичь времен переключения около 100 нс, однако они не могут обеспечить необходимо высокую ширину полосы частот передачи данных, что потенциально вполне реализуемо в оптических системах. Наоборот, хотя перекрестные сети на основе ПМС в принципе способны обеспечить ширину полосы частот передачи данных более 1 ГГц, ожидаемые для них максимальные времена переключения ограничены микросекундными временами. Обсуждавшиеся здесь волоконно-оптические матрицы исключают необходимость использовать ПМС, тем самым сохраняя лучшие черты как оптических, так и электронных перекрестных переключателей. Такие волоконно-оптические матрицы потенциально способны обеспечить и большую ширину полосы частот передачи данных, и малые времена переключения.

В табл. 9.2 представлена таблица истинности для ПЛМ, реализующей все логические комбинации, необходимые для обслуживания переключателя размером 4×4 . Из нее следует, что использование управляющих сигналов позволяет осуществлять выбор любого из вариантов соединений. Рисунок 9.9, а иллюстрирует установление управляющих сигналов для двух вариантов соединений. Хотя коэффициенты объединения по входу и разветвления по выходу, номинально-необходимые для реализации табл. 9.2, можно определить описанным выше способом исходя из числа декодированных входных сигналов и числа термов произведения, но фактически эти коэффициенты могут быть значительно ниже. Например, максимальный коэффи-

коэффициент объединения по входу для волокна вычисляется путем суммирования всех нулевых или единичных входов вдоль заданной строки в табл. 9.2. В этом случае значение максимального коэффициента объединения по входу для волокна равно сумме числа управляющих разрядов 2 и числа разрядов переменных 1, что составляет 3. Максимальный коэффициент разветвления по выходу вычисляется путем суммирования всех аналогичных цифр вдоль заданного столбца. Эта процедура отличается от вычисления коэффициента объединения по входу, поскольку каждый столбец содержит 1-разрядный декодер, разделяющий единицы от нулей. Для данного примера максимальный коэффициент разветвления по выходу для волокна равен 4. Аналогичным образом можно определить требования, предъявляемые к перекрестным схемам размером 32×32 и 128×128 . Эти результаты сведены в табл. 9.3. Из таблицы следует, что перекрестный переключатель 32×32 имеет максимальный коэффициент объединения по входу, равный 6, и максимальный коэффициент разветвления по выходу, равный 32. Обобщая данные табл. 9.3, можно сделать заключение, что перекрестный переключатель имеет коэффициент объединения по входу, равный $n+1$, где $N=2^n$, и величину коэффициента разветвления по выходу, равную N . Полное число входных каналов равно числу управляющих каналов $n \times N$. К этому следует

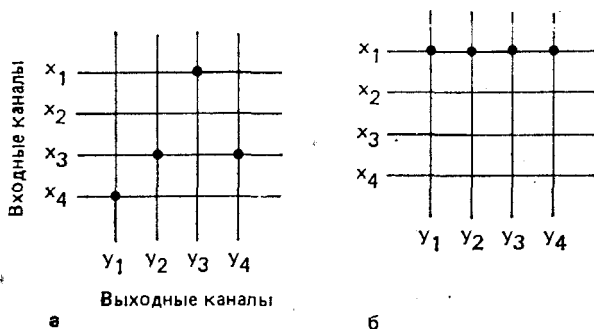


Рис. 9.9. а — управляющие входные сигналы для произвольно выбранного состояния перекрестного переключателя.

Выходной канал

y_1
 y_2
 y_3
 y_4

Управляющие входы

$c_7=1, c_8=1$
 $c_5=1, c_6=0$
 $c_3=0, c_4=0$
 $c_1=1, c_2=0$

б — управляющие входные сигналы для другого произвольно выбранного состояния переключателя.

Выходной канал

y_1
 y_2
 y_3
 y_4

Управляющие входы

$c_7=0, c_8=0$
 $c_5=0, c_6=0$
 $c_3=0, c_4=0$
 $c_1=0, c_2=0$

Таблица 9.3. Сравнение требуемых характеристик перекрестных схем ОПЛМ для различной степени сложности перекрестных соединений

	Размер перекрестной схемы		
	4×4	32×32	128×128
Полное число входов	$2 \times 4 + 4 = 12$	$5 \times 32 + 32 = 192$	$7 \times 128 + 128 = 1024$
Число термов произведения	$4 \times 4 = 16$	$32 \times 32 = 1024$	$128 \times 128 = 16384$
Коэффициент объединения по входу для волокна (максимальный)	$2 + 1 = 3$	$5 + 1 = 6$	$7 + 1 = 8$
Коэффициент разветвления по выходу для волокна (максимальный)	$4 \times 1 = 4$	$32 \times 1 = 32$	$128 \times 1 = 128$

прибавить N каналов передачи данных, в то время как полное число выходных каналов равно N . Для таких устройств полное число произведений равняется $N \times N$. Поскольку управляющие переменные могут переключаться с той же скоростью, что и данные, то скорость переключения перекрестных логических переключателей потенциально является достаточно высокой.

9.4.2. Динамически программируемые логические матрицы

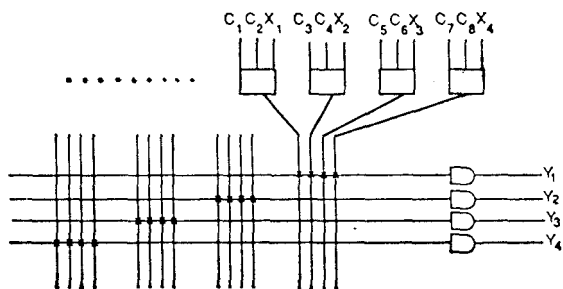
В универсальных перекрестных переключателях конфликтные ситуации предотвращаются за счет того, что в них выполняется только одно соединение на столбец. В противоположность этому ПЛМ допускает любое произвольное число соединений. В принципе логический эквивалент динамической ПЛМ размером $N + N$ может быть осуществлен с помощью ОПЛМ способом, аналогичным описанному выше для перекрестных переключателей, а именно простым соединением N таких устройств. Этот подход, однако, не является особенно привлекательным, поскольку размер и сложность получающейся структуры могут оказаться неприемлемыми. Использование ПМС предоставляет другой способ, но он не способен обеспечить скорость переключения системы, сравнимую с возможностями ОПЛМ. Одним из возможных решений является добавление к ОПЛМ каскада предобработки с другим уровнем логики.

На рис. 9.10, а, б изображена комбинация из ОПЛМ, соединенной со структурой, выполняющей предобработку данных, что в сумме обеспечивает возможность эффективного динамического сброса данных [17]. Структура предобработки данных основывается на простых и повторяющихся логических блоках, которые в сущности являются вариантом соединения типа ячейка—ячейка, используемым в непредставляющем хлопот

случае, когда используют 0 либо 1. Комбинация ОПЛМ и логической предобработки полностью эквивалентна подходу, основанному на использовании ПМС, однако операция перестраивания сети может быть выполнена за один тактовый цикл. Для N каналов данных указанный метод позволяет получить N выходных каналов. При добавлении узла локальной памяти устройство может быть построено возвращено в исходное состояние. Для этого требуется приблизительно $2 \times N$ управляющих каналов. Это может быть сделано в значительной мере подобно тому, как и в случае с обычным ПМС.

9.5. Выводы

Анализ возможностей масштабирования, проведенный в разд. 9.2, показывает, что большие значения ширины полосы частот и коэффициентов разветвления по выходу и объединения по входу волоконно-оптических ПЛМ обеспечивают работу в режимах, недоступных для электронных логических устройств. Принципиальным преимуществом ОПЛМ является способность рассеивать мощности на значительных площадях. В разд. 9.2



а

C_1	C_2	x_1	f_1
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

б

Рис. 9.10. а — вариант соединения ПЛМ с устройством логической предобработки данных в динамически программируемой ОПЛМ. б — таблица истинности для повторяющихся логических блоков, показанных в части а.

также показано, что и произведение числа межэлементных соединений на полосу частот, и произведение коэффициентов разветвления по выходу и объединения по входу на ширину полосы частот являются мерой производительности. Эти величины также пропорциональны произведению мощности, ширины полосы частот и чувствительности детектора. Таким образом, при постоянной ширине полосы частот мощность, необходимая для достижения определенной производительности, оказывается прямо пропорциональной произведению коэффициентов разветвления по выходу и объединения по входу.

В разд. 9.3 сделан вывод о том, что для определенного уровня функциональной сложности использование декодеров высших порядков приводит к монотонному снижению полного числа термов произведения, или, что эквивалентно, сложности вычислений. Напротив, производительность, а также мощность и энергия, требуемые для конкретных вычислений, имеют отчетливый минимум. Для эффективной работы, таким образом, требуется оптимальное согласование типа декодера и сложности функции, которую надо вычислить. Для многих задач степень сложности оптимального декодера может быть настолько велика, что не может быть реализована на основе ОПЛМ. К счастью, существует сравнительно простой способ синтеза генератора функций на основе электронных чипов ЗУПВ или ПЛМ, дающих возможность обойтись одним выходным каналом, который для определенных термов произведений эквивалентен целой группе выходных каналов декодера. Повторяя процедуру для каждого из входов минимизированной таблицы истинности и соответственно после нахождения И и ИЛИ входов, можно реализовать очень большие таблицы истинности, используя при этом сравнительно малое число электронных чипов предварительной обработки, а также располагая сравнительно скромными возможностями ОПЛМ во втором каскаде. Суммарный эффект применения этого метода работы с таблицей истинности состоит в существенном снижении требований, предъявляемых к коэффициентам объединения по входу и разветвления по выходу самой ОПЛМ. Тем самым удается обойти необходимость применения в реальных устройствах сложных декодеров. Путем подходящей замены (в функциональных генераторах) коэффициентов разветвления по выходу и объединения по входу для электронных устройств на коэффициенты для оптических ОПЛМ удается сдвинуть оптимальную область порядка (сложности) декодера в сторону значительно больших величин. В качестве примера рассматривается ситуация, в которой для многих функций сорока переменных процедура синтеза, проводимая с помощью двух 20-разрядных декодеров, дает менее 1000 минимизированных произведений. В этих условиях в большинстве случаев оказывается достаточным использовать каскады

предварительной обработки, состоящие из приемлемого числа чипов ЗУПВ, работающих с 20 переменными (1 Мбит). По мере того как становятся доступными чипы предварительной обработки с большими плотностями, можно ожидать возрастания потенциальных возможностей ОПЛМ в геометрической прогрессии. Например, использование двух 24-разрядных декодеров должно обеспечить возможность получения большого числа функций 48 переменных на основе ОПЛМ со скромными возможностями и приемлемым числом чипов предварительной обработки.

Проблема синтеза представляет собой совсем другую задачу. Процесс логической минимизации требует столь интенсивных вычислений, что для проведения минимизации в задачах с более чем 30 переменными требуются суперкомпьютеры. Методика ОПЛМ тогда оказывается в трудной ситуации, будучи в настоящее время неспособной реализовать свои возможности в полной мере отнюдь не из-за ограниченных возможностей технологии ОПЛМ, но скорее из-за ограниченных возможностей технологии электронных вычислительных устройств. Для некоторых функций, как, например, в случае перекрестных переключателей или динамических программируемых логических матриц, их минимизированные варианты являются настолько простыми, что они могут непосредственно быть получены при проверке представленных данных. В этих ситуациях можно ожидать, что ОПЛМ будут способны работать в режимах, использующих сотни переменных входных сигналов. Следует заметить, что униполярные пороговые функции с 1-разрядными весовыми коэффициентами также могут быть синтезированы сравнительно просто вследствие симметрии в минимизированных термах произведения и из-за взаимнооднозначного соответствия этих термов с биномиальными коэффициентами.

Недавно проведенные эксперименты указывают, что реально достижимы большие значения коэффициентов разветвления по выходу и объединения по входу и производительности вычисления [32]. В данном разделе были подчеркнуты потенциальные возможности волоконно-оптических ПЛМ с точки зрения производительности. На практике такие системы могут оказаться важными даже в ситуациях, требующих сравнительно малого числа операций И и ИЛИ. Одна из таких ситуаций может возникнуть, например, в сложных логических цепях, построенных с помощью соединений многих чипов. Когда длина соединений между электронными логическими чипами превышает 6 дюймов, маленькие волоконно-оптические ПЛМ могут оказаться просто незаменимыми. Гибридное второе поколение архитектур этого типа детально описано, например, в [16]. В другом предельном случае, если размер ОПЛМ существенно увеличен, то должна уменьшиться ширина полосы частот. При

ширине полосы частот менее 100 МГц может оказаться полезным ввести в структуру ОПЛМ электронное запоминающее устройство. При указанных обстоятельствах достаточно обоснованными представляются ожидания, что специализированные архитектуры, использующие группы таких блоков, могут быть разработаны для того, чтобы обеспечить мощные универсальные вычислительные ресурсы.

Проблематичными являются вопросы о том, будут ли эти системы успешно развиваться как одиночные устройства или как процессоры и будут ли они называться оптическими или электронными; тем не менее современные тенденции дают основания полагать, что системы межэлементных соединений оптоэлектронных устройств будут играть все более важную роль в развитии будущих компьютерных систем.

Часть IV

СИМВОЛЬНЫЕ ВЫЧИСЛЕНИЯ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Глава 10

ОПТИКА И СИМВОЛЬНЫЕ ВЫЧИСЛЕНИЯ

Дж. А. Нефф, Отдел электроники, Агентство оборонных исследовательских проектов, Арлингтон, шт. Вирджиния

Б. Дж. Кашнер, Отделение информационных технологий БДМ корпорэйшн, Арлингтон, шт. Вирджиния

10.1. Введение

В настоящее время наблюдается очень быстрое развитие вычислительных систем, обладающих интеллектуальными возможностями. Наряду с этим происходит быстрый рост квалификации разработчиков вычислительной техники и значительной части пользователей компьютерных систем. В целом под ИИ понимают всю совокупность аппаратных средств и человеческих навыков, однако очень трудно дать точное определение тому, что подразумевается под понятием ИИ, поскольку интеллект — это относительное понятие, которое не может быть точно измерено или определено. Некоторые элементы интеллекта присутствовали в компьютерах с самого начала, хотя сам термин ИИ вызывает в воображении образ исключительно совершенной компьютерной техники. Например, наличие памяти обычно связывают с интеллектом, однако даже ранние варианты компьютеров включали в себя некоторые виды памяти. Рабочее определение ИИ может быть кратко описано как наделение компьютеров теми качествами, которые приближают их к процессам мышления человека и при этом заметно отличаются от способа работы традиционных компьютеров.

Важной характеристикой систем ИИ является ввод и использование новых знаний в каждой операции вычислений. Специалисты по компьютерам затратили большие усилия на то, чтобы выяснить, как наилучшим образом следует представлять знания в вычислительных машинах. Были разработаны многочисленные методики создания, накопления и обработки символьных структур данных. Эти структуры, или элементы знания, могут быть использованы для описания объектов, событий, знаний о тех или иных действиях и знаний о том, что уже известно (или метазнания). В целом такой набор символьных структур называют базой знаний системы ИИ.

Почему оптическая обработка информации выглядит привлекательной для выполнения символьных операций или вычислений? Данная глава предназначена дать достаточно подроб-

ный ответ на поставленный вопрос, но рассмотрение ограничено лишь вопросами, связанными с производительностью вычислений и совместимостью разных режимов работы. Требования к производительности вычислений являются следствием сравнения характеристик производительности электронных цифровых компьютеров и машин, использующих язык Лисп, иначе Лисп-машин. Буквально Лисп означает «обработка списков», данный язык символьных вычислений широко распространен в США в научной среде, связанной с проблемами ИИ. На рис. 10.1 представлена производительность вычислений ряда систем ИИ в зависимости от скорости выполнения операций [1, 2]. В качестве единицы измерения интеллекта используется лисп, или число логических выводов в секунду, поскольку в настоящее время для систем ИИ не существует функционального эквивалентного IQ-теста¹⁾. Из данных, представленных на рисунке, видно, что Лисп-машины имеют весьма низкое быстродействие по сравнению с современным поколением «суперкомпьютеров» и мультимикропроцессоров.

Потребность в резком увеличении скорости обработки данных в системах ИИ, в настоящее время ограниченной предель-

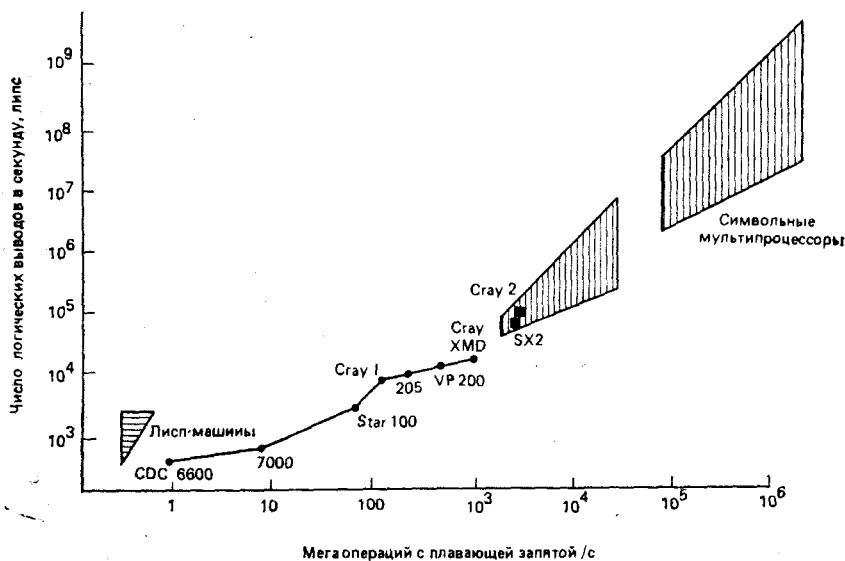


Рис. 10.1. Тенденции в цифровых и аналоговых вычислительных технологиях (по данным ИИЭР).

¹⁾ IQ-тест — система тестов, направленная на определение коэффициента интеллектуальности, показателя умственного развития, уровня имеющихся знаний. — Прим. перев.

ными возможностями современных архитектур монопроцессорных систем, является основным стимулом для исследования схем параллельной обработки, что направлено на повышение быстродействия и расширения круга практических приложений компьютерных систем ИИ. Необходимость использования методов параллельной обработки информации указывает на потенциальный синергизм между символьной и оптической видами обработки.

Представляется интересным тот факт, что цифровые суперкомпьютеры выполняют функции ИИ с большей скоростью, чем специальные машины ИИ. В частности, это обстоятельство можно понимать в том смысле, что именно повышение быстродействия является важным и универсальным средством в преодолении «узких» мест в задачах ИИ. Это также можно интерпретировать и таким образом, что архитектуры, сконструированные для увеличения пропускной способности цифровых систем, будут полезны для символьных систем, и наоборот. Задача описания характерных свойств вычислительных структур, обладающих необходимыми функциональными возможностями, в настоящее время представляет собой глобальную тему для рассмотрения и требует использования практически всех областей компьютерной техники и математики.

Второй фактор, потенциально связывающий оптическую и символьную обработки данных, связан с проблемой совместности операций, и вытекает из необходимости выполнять операции корреляции, поиска и сравнения для различных символьных данных. При этом многие из этих операций не требуют высокой точности вычислений. А при использовании оптики в символьных вычислениях, по-видимому, удастся избежать традиционной «ахиллесовой пяты» оптических вычислений — трудности достижения точности вычислений более нескольких битов. В дополнение к этому наличие тесной связи между символьными вычислениями и вычислениями корреляционных функций (и изоморфных им) может предоставить великолепную возможность улучшения характеристик символьных вычислительных систем с помощью оптических корреляторов.

В разд. 10.2 описаны основные свойства символьных вычислений и проводится сравнение их с распространенными методами цифровых вычислений. Изложены наиболее важные понятия символьных вычислений и ИИ. Затем рассматривается круг проблем, которые, по-видимому, будут встречаться в будущем на пути реализации этих возможностей. В этих разделах непосредственно не затрагиваются вопросы оптических символьных вычислений, но содержащаяся там информация, по-видимому, будет важной для читателей с точки зрения разъяснения сути синергизма, существующего между оптическими и

символьными вычислениями. По замыслу авторов указанные разделы гл. 10 должны выполнить роль введения в проблему ИИ; из-за необъятности проблемы обсуждение придется ограничить только основными аспектами. В разд. 10.4 изложены основные представления об оптических символьных вычислениях. Описаны фундаментальные концепции архитектуры; основной упор сделан на то, что они существенно отличаются от более традиционных архитектур компьютеров. Вслед за этим изложены концепции авторов данной главы относительно того, как оптические методы могли бы улучшить характеристики символьных процессоров.

Стараясь охватить весьма широкий круг проблем, авторы сочли целесообразным сделать введение к каждой из основных дисциплин, составляющих область символьных вычислений. Однако во многих случаях, чтобы обеспечить пропорциональное освещение разных вопросов, потребовалось пожертвовать многими деталями конкретных вопросов. В ряде случаев материал просто выходил за рамки данной книги, и за дополнительной информацией, интересующей читателя, следует обратиться, например, к такому литературному источнику, как [3], представляющему собой трехтомный трактат и содержащему прекрасный обзор по методам ИИ.

10.2. Что такое искусственный интеллект?

Для того чтобы реализовать компьютеры, способные подражать процессам человеческого мышления, в отличие от существующих сегодня цифровых компьютеров, в дополнение к битам с цифровой информацией следует предусмотреть резервные биты для записи определенных символов. Например, кроме того что компьютер должен дать ответ на вопрос: «Каким путем я должен достичь поставленной задачи, располагая известной мне информацией об окружающих меня условиях и моих собственных способностях?», он должен быть способен выполнять арифметические задания типа «Сколько будет 2 плюс 4?». На практике люди вполне стандартным путем находят стандартные решения указанных задач, но пока не удалось создать цифровых компьютеров, выполняющих такие задания. Прежде всего само кодирование поставленного вопроса при вводе в компьютер создает значительную проблему. Во-вторых, если задание и информация об альтернативном пути решения определяются визуальной сценой (т. е. управление процессом решения задачи осуществляется на основе визуальной информации), то использование символьных операций существенно расширяет возможности системы по распознаванию и пониманию. Наконец, процесс принятия решения, по всей вероятности, потребует знания об окружающем мире, а это также основывается на ис-

пользовании аналоговых логических структур. Как показано на рис. 10.2, сбор знаний может происходить тремя различными путями. Во-первых, знание может быть введено в компьютер программистом, работающим совместно с инженером по знаниям. В случае экспертных систем (обсуждаемых в разд. 10.3.5) инженер по знаниям получает знания или проводит их отбор с помощью эксперта и передает их программисту, который загружает знания их в компьютер. Данный процесс называется инженерией знания. Во-вторых, можно ожидать, что уже в ближайшем будущем сбор данных все в большей степени будет осуществляться автоматически, с помощью датчиков. Популярность таких систем будет возрастать по мере создания совершенных систем распознавания речи и систем технического зрения (обсуждаемых в разд. 10.3.2 и 10.3.3). Сбор запрограммированных и считанных данных может выполняться чисто механически, но наибольшая часть информации обычно определенным образом классифицируется, чтобы облегчить в дальнейшем процесс поиска. Классификация — это процесс, осуществляемый при проведении поисков в базе данных и представляющий собой установление связей между каждым входным сигналом и связанными между собой элементами. Цель данного процесса состоит в образовании классов, предназначенных в основном для идентификации данных, относящихся к интересующему вопросу. Классификация также часто рассматривается как процесс организации связей и выделения структур. Если известно, что после сбора заданного объема информации она будет сгруппирована с помощью определенных объектов, имеющих в базе знаний, тогда между двумя из объектов устанавливается связь (обозначаемая линией на схеме). Если нужно одновременно использовать несколько объектов, то они объединяются в более крупную структуру. Читатель будет хорошо представлять и понимать процесс классификации данных после того, как ознакомится с проблемой поиска и представления знаний.

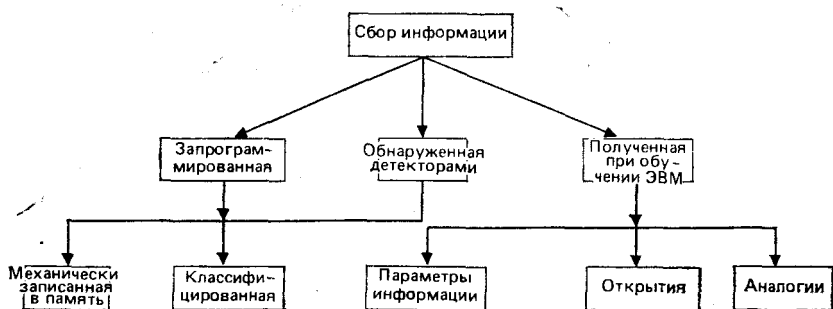


Рис. 10.2. Широко распространенные способы сбора информации.

10.2.1. Характеристики знаний

Система, пригодная для работы со знаниями, должна обладать способностью накапливать дополнительные знания, извлекать соответствующую информацию из базы знаний, проводить рассуждения исходя из конкретной проблемы и информации. Следует признать, что возможно и другое определение понятия знаний, опирающееся на другой набор признаков. Однако авторы данной главы уверены в том, что предложенный ими набор признаков является наиболее удобным, и именно он будет использоваться в дальнейшей дискуссии. Тем не менее данное определение страдает тем же недостатком, что и любая попытка аккуратно проанализировать и классифицировать сложное явление — оно не способно установить степень независимости признаков. Это станет очевидным при обсуждении процедур сбора данных, в котором способность проводить рассуждения (третий признак) описывается через метод сбора данных (первый признак), т. е. представляет собой не что иное, как процедуру обучения.

Сами методики обучения ЭВМ еще пока находятся на начальной стадии развития, но тем не менее можно выделить три направления, которые успешно развивались в последнее время. Они включают процедуры корректировки параметров, проведения рассуждений по аналогам [3, 4]. Изменение параметров и стимулов является обычной научной методикой обучения. Две важные области применения этого метода к задачам ИИ состоят в изменении классификационных параметров, используемых при сборе знаний (изменении классов, на которые разбивают все множество объектов), и в корректировке параметров эвристической функции с целью улучшения процедуры решения задач (последняя методика будет обсуждаться в разд. 10.2.3). Факт обнаружения нужной информации (знания) обычно влечет за собой решение проблемы, и, следовательно, основывается на проведении рассуждений. Процесс проведения рассуждений также лежит в основе обучения по аналогам, которое включает в себя получение недостающей информации о некотором объекте, если известно, что неполностью определенный объект «подобен» известному объекту. Представление знаний с помощью фреймов и сценариев (рассматриваемых в разд. 10.2.3) облегчает проведение рассуждений по аналогии, так как можно непосредственно «пропустить» имеющиеся признаки, например, через два фрейма или документа, являющихся, согласно ожиданиям, близкими друг другу.

Для систем ИИ выборка данных является весьма сложной проблемой. Это обусловлено несколькими причинами, одна из которых связана с большим размером баз данных. Эти базы данных содержат не только набор данных, относящихся к про-

блемной области данной системы, но содержат также и правила, позволяющие проводить достаточно интеллектуальные процедуры с данными. Одна из общих методик выборки данных использует наборы индексов, которыми маркируют данные по одному или большему числу признаков. Например, голографическая линза могла бы быть индексирована по таким признакам, как «оптическое устройство», «дифракционная решетка», «согласующее устройство», «узкополосное устройство», «поверхностное (тонкопленочное) устройство». Язык программирования Лисп, являющийся наиболее широко распространенным языком в областях, связанных с ИИ, облегчает присвоение признаков объектам в том плане, что он формирует вокруг определенного символа списки связанных с ним символов. Приведенный выше пример мог бы быть закодирован в языке Лисп с помощью «списка свойств» следующим образом:

(голографическилинзы оптические дифракционныерешетки
согласующие узкополосные поверхностные).

Две другие схемы выборки данных основаны на сопоставлении с образцом и использовании контекста. Схемы машинного сопоставления с образцом осуществляют выбор данных, соответствующих определенным шаблонам, связанным с определенными категориями данных. Более совершенная схема выборки может быть использована при контекстуальном накоплении данных, когда данные извлекаются в соответствии со своими значениями. В качестве примера машинного сопоставления с образцом рассмотрим базу данных, содержащую следующие списки:

(источниксвета лазер гелийнеоновый длиनावолны(x)...)

(источниксвета Nd—АИГ лазер...)

(источниксвета лазер полупроводниковый...)

(источниксвета дуговаялампа ртутная...)

(источниксвета дуговаялампа ксеноновая...)

(источникпитания 110Ввход 12Ввыход...)

Запрос на поиск этих представляющих источники света элементов выглядит как

(источниксвета ?x)

А чтобы найти входы к объектам, связанным с лазерами, образец для сопоставления элементов должен быть изменен следующим образом:

(источниксвета лазер ?x)

Все эти схемы выборки данных значительно отличаются от схем, используемых в цифровых компьютерах, которые накапливают данные в соответствии с их адресом в памяти. Цифровая информация, являясь по существу подмножеством символьной информации, может накапливаться и выбираться с помощью упомянутых выше схем, хотя и не столь эффективно. Например, простой список (+ 2 4 9) связывает цифровые символы 2, 4 и

9 операцией суммирования, а вложенные списки $(+ (*3 4) (*6 3))$ связывают 3 и 4 с одной операцией умножения, 6 и 3 с другой операцией умножения, а 12 и 18 — операцией сложения. В этом примере используется основная форма Лисп-нотации, в которой первый элемент списка представляет операцию, которую надо выполнить, в то время как остальные элементы являются аргументами, над которыми выполняется указанная операция.

В любой дискуссии по проблеме поиска данных следует переходить от средств распознавания релевантных данных в базе знаний к обсуждению того, как понять такой набор или наборы данных, которые приводят к выполнению поставленной цели. Однако обсуждение процедур поиска будет отложено до рассмотрения представления знаний, поскольку эти два вопроса тесно связаны между собой; знания обычно представляют в таком виде, который облегчает сам процесс поиска.

Проведение рассуждений, являющееся третьим атрибутом интеллектуальных систем, работающих со знаниями, необходимо тогда, когда система нуждается в информации, которая не может быть извлечена из базы знаний непосредственно. Системы ИИ могут осуществлять обмен данными между большими базами знаний и сложными процедурами проведения рассуждений. Такие системы должны или обладать высоким уровнем возможностей проведения рассуждений, или быть способными накапливать и выбирать всю релевантную информацию; при этом система, которая затрагивает слишком много времени на поиск стратегии проведения рассуждений, не обладает достаточными знаниями.

Процесс проведения рассуждений можно рассматривать как движение в пространстве состояний, где состояния представляют все возможные ситуации, и движение осуществляется из начального состояния (состояний) через текущее состояние (состояния) в заданное конечное состояние (состояния). Проведение рассуждений при решении практических проблем обычно требует прохождения через многие промежуточные состояния. Допустимые переходы между состояниями определяются либо правилами, задаваемыми утверждениями типа «если...тогда», либо с помощью организации связей между фактами, для обозначения которых используют так называемые направленные графы. В следующем разделе при обсуждении вопросов представления знаний будут изложены различные методики, используемые для определения состояний и переходов между состояниями. Классическим примером реализации концепции пространства состояний может служить игра в шахматы, в которой начальное состояние являлось бы отправной точкой для всех фигур, а состоянием цели была бы любая конфигурация фигур, в которой все возможные перемещения противника были бы заблокированными, и король находился бы под ударом. Перехо-

ды между состояниями регламентировались бы правилами, указывающими допустимые перемещения каждой из фигур в каждом положении (каждое состояние имело бы различные допустимые перемещения, зависящие от расположения других фигур и близости фигур к краю доски).

Факт пребывания в данном состоянии наиболее удобно описать с помощью решателя задачи, имеющего большое число возможных переходов в другие состояния и указывающего путь к решению. Следовательно, операция поиска также играет ведущую роль в процессе проведения рассуждений. В данном случае поиск осуществляется согласно пути (или путям) в пространстве состояний в рамках заданного круга проблем, в то время как при операции выборки данных производится поиск релевантных данных в базе знаний. Как уже указано выше, операции поиска будут обсуждаться после вопросов представлений знаний, рассматриваемых в следующей части раздела.

10.2.2. Представление знаний

Для каждого из обсуждавшихся выше видов знаний принципиальное значение имеет вопрос о том, в каком виде следует представить их в компьютере, чтобы облегчить их взаимодействие между собой и сделать систему более полезной для решения практических задач. Разработаны многочисленные варианты представления знаний, но большинство из них является вариантами или комбинациями следующих четырех: семантические сети, системы продукций, фреймы и логические системы. Семантические сети по своей природе очень разнообразны, но в общем могут быть охарактеризованы как схемы графического представления, в которых вершины графов представляют собой объекты или концепции, а соединения между вершинами — процедуры получения выводов, связанных с соответствующими вершинами. На рис. 10.3 изображена крайне упрощенная семантическая сеть, демонстрирующая получение вывода о том, что в брегговской ячейке наблюдается распространение как поперечных, так и продольных волн. Этот тип знаний часто называют декларативным, так как его часто получают на основе реальных ситуаций, связанных с конкретными знаниями или отношениями.

В своих наиболее простых проявлениях системы продукций представляют собой форму представления знаний в виде множества правил (продукций) «образец/действие» типа «если/тогда». Если образец в правиле является истинным («образец» также называют «условием»), тогда в силу вступает «действие», т. е. состояние вычислительной машины изменяется согласно определенному «действию». Примерами таких правил могут служить следующие:

«Если источник света лазер, то он излучает когерентный свет», и «Если важна низкая стоимость и если не требуется когерентность света, тогда следует использовать светодиод вместо полупроводниковых лазеров».

Системы продукций широко используются в качестве схем представления данных в экспертных системах (рассматриваемых в части разд. 10.2.5). Часто системы продукций называют процедурными знаниями, так как обычно выполнение определенных действий инициируется известными правилами. Следовало бы заметить, что сравнение образов (корреляция) играет важную роль в системах продукций в том плане, что выполнение (запуск) правил основано на совпадении условий правил (частей «если») и состояний задачи, т. е. процесс сравнения определяет истинно или ложно условие.

Представление в виде фреймов использует организацию данных по функциональным группам иерархически связанных значений признаков. Такое представление обладает рядом преимуществ при работе со стереотипными концепциями, такими как показаны на рис. 10.4. Данный фрейм, содержащий знания о двумерных модуляторах света, может быть представлен с помощью одного или нескольких фреймов, использующихся для описания систем оптической обработки информации, устройств ввода/вывода или оптических приборов. В свою очередь обсуждаемый фрейм, описывающий модулятор света, связан с другими, зависящими от него фреймами (связанными с его входами); такие связи на рисунках показывают стрелками. Например, введение в схему обработки признака нелинейного оптического материала привело бы к появлению фрейма с информа-

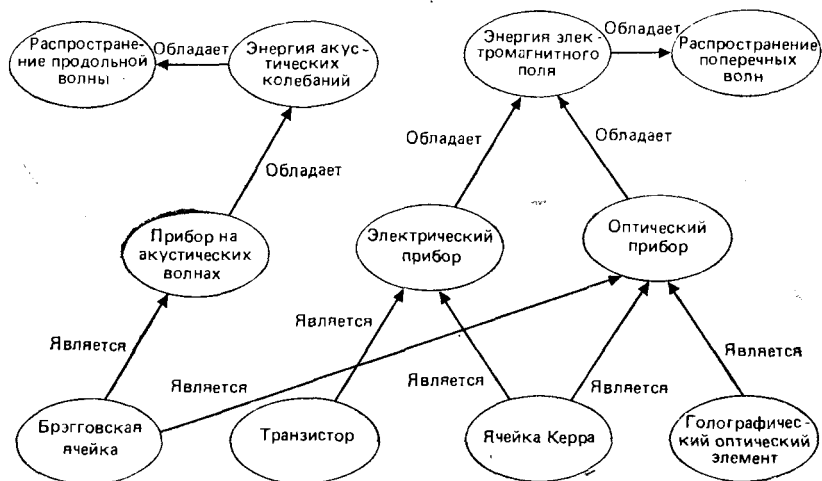


Рис. 10.3. Упрощенный пример семантической сети.

цией о признаках таких материалов. С концепцией фреймов связана концепция сценариев, использующая набор общих последовательностей выполнения событий, аналогично тому, как во фреймах используют набор связанных между собой объектов и признаков. Например, можно было бы сформулировать сценарий изготовления модулятора света из его отдельных компонентов.

В логических схемах представления знаний с помощью синтаксических операций с формулами стараются создать особые конструкции из имеющихся знаний, позволяющие сделать заключения, исходя из предпосылок ИСТИНА или ЛОЖЬ. В этих целях на основе разработок в области исчисления высказываний был получен формализм, известный как исчисление предикатов. В целом логическое представление является весьма привлекательным для систем, где часто требуется расширять базу знаний, например при доказательстве теорем. Способность расширять базу знаний основана на методе математической дедукции, позволяющем получать новые факты из уже имеющихся.

Отношения, существующие в реальном мире, выражаются в виде предикатов и их аргументов. Предикаты представляют собой отношения между предметами, а символы предикатов используются для обозначения этих отношений. Предикат, применяемый к своему аргументу (аргументам), возвращает в качестве результата либо «истина», либо «ложь». В семантической сети, показанной на рис. 10.3, связь «является» изображена между вершиной, описывающей ячейку Керра и вершиной, описывающей оптическое устройство. В логическом представлении она была бы представлена предикатом ЯВЛЯЕТСЯОПТИЧЕСКИМУСТРОЙСТВОМ. Запрос к базе знаний в виде предикатного выражения ЯВЛЯЕТСЯОПТИЧЕСКИМУСТРОЙСТВОМ (ЯЧЕЙКАКЕРРА) получил бы ответ ИСТИННО, в то время как запрос ЯВЛЯЕТСЯОПТИЧЕСКИМУСТРОЙСТВОМ (МАГ-

Двумерный пространственный модулятор света

Характерные свойства двумерных ПМС, не приводящие к возникновению дополнительных фреймов

Фотопроводник:	Кремний
Устройства считывания/записи:	Оптические
Нелинейные оптические материалы:	BSO (бисульфит свинца)
Доступные источники:	Лазерный диод
Частота следования фреймов:	1 кГц
Число ячеек:	$10^3 \times 10^3$

Рис. 10.4. Фрейм знания, характеризующего двумерный ПМС.

НИТ) получил бы ответ ЛОЖНО. В качестве других примеров предикатов можно привести (записывая для общности с переменными в качестве аргументов) такие варианты:

ЖЕНЩИНА (x) ИСТИННО если x является женщиной

РАВНО (x, y) ИСТИННО если $x=y$

ПРЕЗИДЕНТ (x, y) ИСТИННО если x является президентом y

В исчислении предикатов возможно определенным образом группировать предикаты с их аргументами (но не с утверждениями типа «здесь светло»), используя в этих целях такие логические связки исчисления высказываний, как И(\wedge), ИЛИ(\vee), НЕ(\sim), РАВНО($=$) и СЛЕДУЕТ(\rightarrow). В дополнение к этому требуется ввести квантификаторы, задающие области определения переменных. Например, утверждение РАВНО (x, y) могло бы означать любое из следующих четырех соотношений: все x равны всем y , заданный x равен заданному y , все x равны некоторому значению y , или некоторые значения x равны любому значению y . Универсальный квантификатор \forall обозначает, что значения переменных x должны быть рассмотрены, в то время как уже существующий квантификатор \exists означает, что должны быть рассмотрены лишь некоторые определенные значения x . Например, если истинность выражения РАВНО(x, y) должна соблюдаться на всех значениях x , равных всем значениям y , тогда можно было бы записать это в виде $\forall x \forall y \text{РАВНО}(x, y)$. Если бы ситуация была такова, что утверждение должно быть ИСТИННО, если конкретные значения x равны определенным значениям y , тогда следовало бы записать $\exists x y \text{РАВНО}(x, y)$. Примерами выражений, записанных в представлении предикатов, являются:

Свет является когерентным, если он испускается лазером.

$\forall x (\text{ИЗЛУЧЕНИЕ ЛАЗЕРА}(x) \supset \text{КОГЕРЕНТНЫЙ СВЕТ}(x))$

Излучение некоторых лазеров является видимым.

$\exists x (\text{ИЗЛУЧЕНИЕ ЛАЗЕРА}(x) \wedge \text{ВИДИМЫЙ СВЕТ}(x))$

Прежде чем вводить представления о последнем из основных элементов логики, т. е. о функциях, целесообразно сделать обзор ранее введенных элементов. Они кратко описаны на рис. 10.5. Следует заметить, что традиционно переменные обозначают строчными буквами, а константы и предикаты — прописными.

Предикаты являются несколько ограниченными по содержанию средствами, поскольку их оценки дают только значения ИСТИННО или ЛОЖНО. Например, предикаты ЖЕНЩИНА (ЖЕНА), ЖЕНЩИНА (ФЛОРЕНСНАЙТИНГЕЙЛ) и ПРЕЗИДЕНТ (СОЕДИНЕННЫХ ШТАТОВ, ДЖОРДЖ ВАШИНГТОН) получает ответ ИСТИННО, если в последнем из приме-

ров ПРЕЗИДЕНТ(x, y) было присвоено значение, « y является президентом x ». ЖЕНЩИНА (ДЖОРДЖВАШИНГТОН), конечно, получил бы ответ «ЛОЖНО». С другой стороны функции могут давать ответы в виде объектов, которые соответственно используются как аргументы предикатов. Иллюстрацией к данному случаю является функция «длина волны (x)», используемая для поиска числовых значений длины волны, связанной с x ; т. е. длина волны (КРАСНЫЙСВЕТ) «даст ответ 0,6 мкм» и «длина волны (СО2ЛАЗЕР)» даст ответ 10,6 мкм. Примером функции, являющейся частью предиката, являлась бы ВИДИМАЯОБЛАСТЬ (длина волны (СО2ЛАЗЕР)), которая получила бы ответ ЛОЖНО, так как СО2 лазер излучает в ближней инфракрасной области спектра, а не в видимой.

10.2.3. Поиск

Большие размеры баз знаний и пространства состояний, необходимые для решения практических символических задач, обуславливают обязательное применение стратегии поиска либо так называемых релевантных фактов, содержащихся в базе знаний, либо путей решения в пространстве состояний задачи. На рис. 10.6 изображены различные виды поиска, включая чисто случайный поиск и весьма специфический эвристический поиск (эвристика предполагает использование знаний по определенной проблеме с целью придать поиску направленный характер). Хотя на этой упрощенной диаграмме стратегии поиска разделены на конкретные категории, однако фактически такого четкого разграничения провести невозможно. Чем менее направленным является поиск, тем большее время требуется для достижения цели. С другой стороны, чем более целенаправленным

Описание	Примеры	Комментарии
Переменные	x, y	Используется для универсальных аргументов
Константы	ЯЧЕЙКАКЕРРА, МАГНИТ	Специальные наименования, присваиваемые аргументам
Символы предикатов	ЯВЛЯЕТСЯОПТИЧЕСКИМ УСТРОЙСТВОМ ()	Определяются таким образом, чтобы представлять соотношения
Логические связи	$\wedge, \vee,$	Выражения взаимосвязей символов
Кванторы	$\forall, \exists,$	Количественно определяют размеры области определения переменных

Рис. 10.5. Элементы исчисления предикатов.

является поиск, тем более сложными оказываются процедуры управления. В предельном, наиболее сложном случае средства управления включали бы в себя такой объем знаний о процессе поиска, что сама необходимость поиска исключалась бы; т. е. средства управления направили бы решение непосредственно на цель. Поскольку оба крайних случая практически не реальны, то обычно делают попытки найти некоторую компромиссную стратегию, являющуюся наилучшей для конкретной задачи. Фактически многие системы ИИ используют комбинацию схем общего и частного назначения, что позволяет видоизменить общую стратегию поиска цели после его осуществления, согласно изменяющейся структуре пространства состояний.

К методикам поиска относят несколько общих стратегий, которые могли бы выполнять роль более или менее целенаправленных процедур в зависимости от той степени, в которой проблемная область и информация о цели применяются при выборе путей поиска. Некоторые из этих стратегий будут рассматриваться ниже. При этом будут сравниваться между собой древовидная структура и направленный граф, метод обхода дерева «перебором в глубину» и метод «перебора в ширину», а также методика вывода «от фактов к цели» и методика «от цели к фактам».

Процесс поиска может быть осуществлен путем следования либо по структуре направленного графа, изображенного на рис. 10.7, а, либо по соответствующей древовидной структуре, показанной на рис. 10.7, б. Поиск методом направленного графа позволяет вернуться к одной из ранее пройденных вершин графа и продолжить поиск по другому маршруту. С другой стороны, дерево поиска может совершить повторы при реализации стратегий. Например, при поиске устойчивых решений узел «С» может быть пройден повторно (например, через вершины А С F...), если первая попытка (через А В Е С) не позволила достичь цели. Очевидным недостатком древовидной структуры

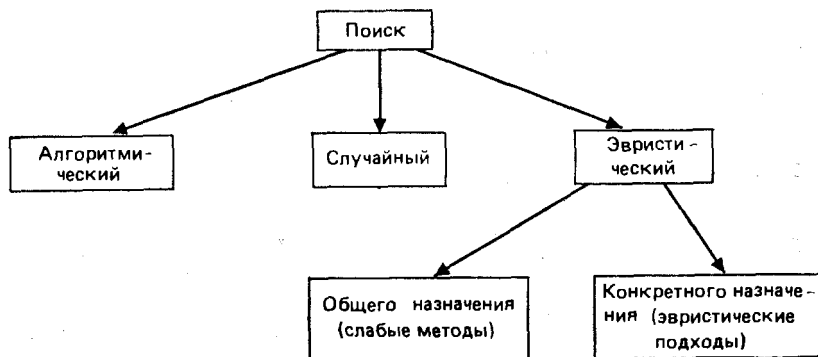
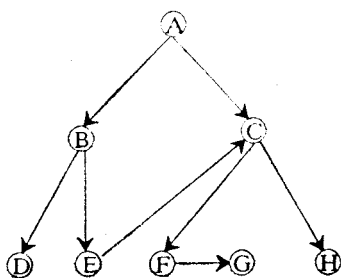


Рис. 10.6. Иерархия стратегий поиска.

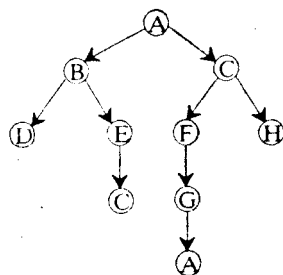
является избыточность числа проходов, а одним из преимуществ — использование намного меньшего объема памяти из-за того, что запоминать предыдущую стратегию в данном случае не требуется. Выбор древовидной структуры или графа часто определяется характером решаемой задачи, т. е. тем, насколько велика вероятность возникновения повторения ситуаций. Ресурсы памяти будут также являться решающим фактором.

Схемы поиска также могут отличаться порядком поиска вершин. Строго последовательный поиск, известный как перебор в глубину, следует заданной стратегии (например, ветви или дереву) до тех пор, пока стратегия не приведет к успеху в достижении цели, или не будет выявлено, что пора прекратить безуспешный поиск. В последнем случае происходит возврат к последней пройденной вершине, имеющей еще непроверенные ветви. Такая процедура поиска экономит ресурсы памяти, поскольку оказавшиеся бесполезными ветви исключаются из рассмотрения, однако, если проблемная область характеризуется длинными процедурами поисков, то перебор в глубину может потребовать слишком больших затрат времени. Приведенный на рис. 10.7, б пример поиска перебором в глубину осуществляется по узлам А В D В E C A C F C ...

С другой стороны, в методе поиска перебором в ширину все состояния, связанные с начальным, проверяются на соответствие поставленной цели, и лишь после этого осуществляется переход на более глубокий уровень дерева или графа. Если цель не была достигнута, тогда эти состояния (или вершины) на первом уровне расширяют, т. е. генерируют и тестируют все возможные связанные с ними состояния. Например, поиск в ширину, осуществляемый по дереву целей, изображенному на рис. 10.7, б, происходит в порядке прохождения вершин А В С D E F H C ... до тех пор, пока не будет достигнуто состояние цели. Данный тип поиска может оказаться очень полезным в



а



б

Рис. 10.7. а — поиск с помощью направленного графа; б — дерево поиска.

системах параллельной обработки, которые будут обсуждаться в разд. 10.4.

До этого места в изложении процедуру решения проблемы и процесс проведения рассуждений представляли как последовательность событий, выполняемую из начального состояния в направлении состояния (состояний) цели. Другой способ, которым иногда пользуются, исходит из заданной цели и выполняется в обратном порядке, при этом стараются удовлетворить начальным условиям. В обратной процедуре проведения рассуждений, известной среди специалистов по ИИ как способ рассуждений «от цели к фактам», сначала находят одно или более состояний, которые могут привести к определению цели, и проверить, достигнуто ли соответствие с начальным состоянием (состояниями). Если нет, то поиск продолжается. Для систем продукций это означает, что в этом случае части «тогда» согласуются между собой и соответственно части «если» запускаются в действие (например, если часть «тогда» используется для нахождения более отдаленной вершины). Конкретно выбор либо процедуры от «цели к фактам», либо процедуры «от фактов к цели» определяется прежде всего двумя факторами — соотношением случаев ветвления с переходом назад по телу программы и случаев ветвления с переходом вперед, а также соотношением числа состояний цели и числа начальных состояний. Другими словами, если сформированное дерево поиска в конкретной проблемной области разветвляется в значительно большей степени при прямой процедуре поиска, чем при обратной процедуре поиска, то в этом случае процедура «от цели к фактам» будет более целесообразной. Если разветвление в обоих направлениях приблизительно одинаково, решающее значение приобретает число состояний. Процедура «от цели к фактам» выглядит более привлекательной при решении задач синтеза сложных объектов, когда существует широкий спектр исходных объектов, на основе которых приводится синтез. Примером служит задача определения того, какие характеристики материала необходимы для оптимального изготовления конкретного устройства. Здесь лучше было бы начать процедуру поиска с состояния цели (требований к устройству), чем с поисков наборов характеристик для всех возможных материалов. С другой стороны, в шахматах проведение рассуждений никогда не может быть проведено «от цели к фактам» вследствие исключительно большого числа путей, ведущих к победе.

Для многих практических задач пространство состояний является настолько большим, что обычно жертвуют поиском гарантированных оптимальных решений в пользу применения специальных стратегий и тактики ограничения процесса поиска и удовлетворяются хорошим решением, отказываясь от наилучшего решения. Процесс поиска в шахматных задачах является

прекрасным примером этого; в противном случае шахматные задачи не могли бы решаться с помощью компьютеров. Как упомянуто выше, варьирование объемами информации может быть использовано для продуманного управления выше обсуждавшимися схемами поиска. Такие методики, как индексация, факторизация, сопоставление с образцом относят к категории эвристических методов общего назначения; однако их роль в ограничении сложных процедур поиска сравнительно невелика, тем более что они часто делают необходимым использование существенно более сложных эвристических методов, по сравнению с указанными выше методиками общего назначения. При индексировании используются заранее установленные схемы присвоения индексов, описывающих состояние задачи, а также задающих процесс запоминания правил, применяемых к каждому из состояний задачи (при этом требуется, чтобы эти правила могли быть связаны с индексом данного состояния). Из области хранения данных об определенном состоянии задачи впоследствии можно вызвать индекс этого состояния, что в свою очередь позволяет вызвать все правила, которые могли быть использованы. Более привлекательной выглядит операция факторизации. Если база знаний может быть разделена на большие множества данных, мало пересекающиеся между собой, то при рассмотрении соответствующей информации о проблемной области целые секции могут быть исключены. Например, если рассматриваемый вопрос связан с диагностикой легочных заболеваний, то для системы не требуется поиск части ее базы знаний, связанной с процедурами при проведении реальных хирургических операций на легком. Ограничение поиска также может быть достигнуто с помощью метода сопоставления с образцом, аналогичного традиционным задачам сравнения, где операция сопоставления используется для проверки того, что были найдены правильная форма, слово и т. д. Сопоставление с образцом часто используется в задачах распознавания речи, понимания естественного языка, распознавания образов; все они будут обсуждаться в разд. 10.3.

Можно ввести количественные коэффициенты, описывающие процесс ограничения поиска, с этой целью используют эвристические или оценочные функции определенных видов. Как показано на рис. 10.6, этот случай относят к классу эвристики частного назначения. Эвристическая функция дает значения весового или стоимостного коэффициента для каждой из вершин, что служит в определенной степени мерой «качества» пути достижения решения вплоть до данной вершины. Для различных проблемных областей будут иметься разные возможности для реализации таких функций, отсюда и появляется термин «частного назначения». Наиболее часто применяются концепции измерения качества пути, основанные на метрических представле-

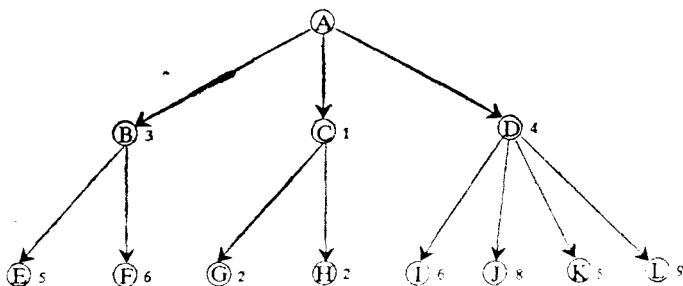


Рис. 10.8. Поиск «наилучшего первого».

ниях расстояния до вершины или степени трудности поиска цели, а также выражаемые разностью путей между текущей вершиной и вершиной цели. Как уже упомянуто выше, в любом случае приходится искать компромисс между возможной экономией времени поиска, достигаемой за счет применения эвристических функций, и затратами времени на вычисления самих функций; т. е. сложные функции могут обеспечить прекрасное управление процессом поиска, но время, необходимое для их вычисления, может быть больше, чем время, которое потребовалось бы для реализации случайной процедуры поиска.

Если задана функция меры качества, описывающая пути поиска решения, то процесс поиска может проводиться по методу «первого наилучшего». Согласно данной методике, в качестве вершин, из которых предстоит расширить поиск цели, выбирают вершины, обладающие наилучшей мерой качества. Например, можно считать, что числа, записанные рядом с каждой из вершин на рис. 10.8, представляют собой значения эвристической функции, так что чем меньше соответствующее число, тем целесообразнее продолжить поиск именно из этой вершины. Соответственно поиск по методу «первого наилучшего» будет происходить следующим образом:

A C G H B D K E F I J L.

10.2.4. Сравнительные характеристики символьных и цифровых вычислений

Надеемся, что читатель уже получил общее представление о том, что составляет суть символьных вычислений. Уровень понимания данного вопроса еще более углубится, если провести сравнение возможностей символьных вычислений и более известных цифровых. На рис. 10.9 для удобства в двух колонках приведен перечень их свойств, так что сравнение можно провести достаточно наглядно. Результаты сравнения обсуждаются ниже.

Логический вывод является аналогом операции с плавающей запятой. В то время как операция с плавающей запятой является основным видом операций при работе с цифровыми символами, процедуры получения логических выводов применяются при работе с гораздо более широким классом символов, используемых в символных вычислениях. Логические выводы получают за счет создания определенных комбинаций элементов знаний или группы объектов с целью получения каких-либо заключений. Для логических выводов, которые получают с помощью силлогизмов (двухступенчатых утверждений «если/тогда»), конкретные способы организации графических связей, сравнения фреймов или методов получения логических выводов существенно зависят от представления использованных знаний (системы продукций, семантической сети, фреймов или соответственно исчисления предикатов первого порядка). Комбинирование и обработка структур символных данных (например, объектов и их признаков) с помощью логических выводов составляют основу для наиболее мощных методик проведения рассуждений в символных вычислениях.

Обладающие выраженной структурой форматы данных, такие как векторы, матрицы и т. д., используемые в цифровых вычислениях, дают возможность создания структур данных, которые могут изменять свой вид при символной обработке. При выполнении таких заданий, как ориентирование на незнакомой территории, различные игры или ведение диалога перед началом выполнения заданий известен только общий вид данных,

<u>Символьные</u>	<u>Цифровые</u>
Логические рассуждения	Операции с плавающей запятой
Динамические структуры данных	Статические структуры данных
Эвристический поиск	Алгоритмический поиск
Допустимы однозначные неполные данные	Необходимость использования точных, полных данных
Лисп, Пролог	Фортран, Паскаль, Кобол,...
Независимость процесса управления и знания по интересующей проблеме	Тесная взаимосвязь управления и имеющейся информации
Способность модифицировать данные на высоком уровне	Модификации на макроуровне
Способность давать разъяснения	Ограниченная возможность дачи разъяснений

Рис. 10.9. Признаки символных и цифровых вычислений.

используемых для планирования маршрута, выполнения ответных действий на действия противника или соответственно содержащихся в произнесенных фразах. Следовательно, символьные процессоры используют списки объектов, соединенных указателями, аналогичными обсуждавшимся в предыдущих разделах данной главы. В отличие от ситуации в цифровых вычислениях, где структуры данных обычно имеют ограниченные размеры, в символьных вычислениях входные данные, обрабатываемые данные и конечная цель их обработки обычно имеют вид фраз, концепций или других символьных структур с неограниченным размером. Более того, динамические структуры данных позволяют символьной машине работать с неполной информацией или приходиться к неопределенным или неоднозначным выводам.

Вопрос о возможности использования неточной или неполной информации демонстрирует читателям суть различий между использованием символьных и цифровых данных. Алгоритмы, использовавшиеся в цифровых компьютерах, обладают слабыми возможностями для получения корректных цифровых выходных данных на основе качественных описаний или неполных цифровых входных данных. Выходные данные зависят от точности входных данных, поскольку все наиболее существенные процедуры сводятся к комбинациям цифровых объектов. С другой стороны, характерным свойством большинства символьных представлений данных является то, что заключения могут быть получены чисто качественно, путем получения определенных соотношений, описывающих функциональные свойства. Это возможно даже тогда, когда входная информация является неполной или в некоторых случаях неточной. Следует заметить, что такие методики находятся на переднем крае сегодняшних исследований символьных вычислений и составляют дисциплину, называемую проведением рассуждений по неполным данным [3].

Вполне очевидно, что для реализации всех возможностей символьных вычислений, таких как использование неточных или неполных знаний при проведении рассуждений, требуются специальные языки программирования. Следует заметить, что языки цифрового программирования, такие как Фортран, в руках умелого программиста могли бы использоваться и в символьных вычислениях. Однако они предназначены для определенных применений и сделали бы аналоговое программирование очень трудной задачей. Язык Лисп и его основные версии Интерлисп и Маклисп создают такую среду программирования, которая существенно облегчает обработку символьных выражений, характеризуемых гибкими структурами данных. Семантические значения объектов изменяются непосредственно при расширении или сокращении переменного списка признаков.

Другой существенной характеристикой языка Лисп является его рекурсивная природа; любая функция языка Лисп может вызывать саму себя, и любая программа может быть определена содержащимися в ней самой понятиями. Такая способность облегчает поиск списков неопределенной длины в процедуре поиска заданных элементов (свойств), содержащихся в списках.

Язык Пролог («программируемая логика») сильно отличается от описанного выше языка Лисп. Язык базируется на правилах системы продукций, и использует методики сравнения образцов, чтобы доказать или опровергнуть какие-либо утверждения программы. В основе языка лежит применение исчисления предикатов для установления соотношений между объектами. Основной операцией в языке Пролог является логическое доказательство некоторых условий или соотношений исходя из набора более простых условий.

На рис. 10.9 указано (шестым по порядку) характерное свойство символьных систем, заключающееся в независимости процедуры управления от знаний, имеющихся по конкретной проблеме. При этом в символьных вычислениях понятие управления распространяется на любой процесс, явно или неявно влияющий на порядок выполнения процедур, осуществляемых с целью решения задачи [34]. Указанное важное свойство ярко проявляется в экспертных системах (обсуждаемых в разд. 10.5.3), где фактическая структура стратегии решения может быть рекурсивно изменена на основе поправок, внесенных в процессе совершенствования программы. В этом отношении ситуация сильно отличается от случая цифровых вычислений, где изменения и даже «условные» ветви в программе вводятся в систему заранее. Именно указанная независимость базы знаний от управляющих процедур позволяет экспертной системе «shell» быть достаточно мощным средством, чтобы применяться сразу в нескольких проблемных областях. Примером также служит тот факт, что хотя программа MYCIN была вначале разработана в целях медицинской диагностики бактериальных инфекций, она тем не менее может быть использована для создания базы знаний в области кристаллографии, предназначенной для специалистов по росту кристаллов. Вместо хранения правил, отражающих симптомы болезней, бактерии и лекарственные средства, база знаний должна содержать правила, относящиеся к методикам измерения кристаллографических структур, и рекомендации по методикам роста кристаллов. В области цифровых методов нельзя, например, взять программу, предназначенную для разработки интегральных схем со сверхвысокой степенью интеграции, и приспособить ее для проектирования линз лишь за счет изменения входных данных.

Способ внесения изменений в символьные программы также

отличается от цифровых методов. Если требуется изменить программу цифрового компьютера, то целая программа или по крайней мере ее часть, куда требуется внести изменения, должна быть выведена из системы, отредактирована и вновь введена в систему; в это же время должен быть перекомпилирован весь (или большая часть) машинный код. При работе с языком Лисп можно модифицировать программу без таких процедур. И программы, и наборы данных в языке Лисп записаны в одном и том же виде, с одним и тем же синтаксисом. В результате этого Лисп-программа может преобразовывать другую программу или базу данных, реализованную на языке Лисп. Она может автоматически исправлять некорректные правила или знания и, кроме того, может выводить вызывающие подозрения правила или знания для контроля оператором, который в свою очередь может вносить изменения в диалоговом режиме без повторной компиляции всей программы. Такие программные средства обладают очень большими возможностями и позволяют оператору организовать специальные «окна» в программе, и таким образом видеть в реальном времени все происходящие изменения в базе знаний или структуре правил. В настоящее время такая форма программного обеспечения не доступна цифровым компьютерам, в которых используются специальные отладочные циклы, требующие существенных затрат времени на поиск ошибок, повторную запись команд, компиляцию, прогоны, поиск новых ошибок, повторную перезапись, компиляцию, прогон и т. д.

Поскольку методы внесения изменений в символьных программах отличаются от цифровых программ, то вполне понятно, что мощность символьных вычислений может также быть использована на стадии отладки. Известно, что в цифровой компьютер возможно встроить программу, позволяющую во время выполнения программы печатать сообщения о тех путях, которые были выбраны в ситуациях «если/тогда». В символьном компьютере, однако, машина может не только указать оператору выбранный маршрут достижения цели, но также сообщить, почему она выбрала каждый из путей. Эта особенность является полезной по нескольким причинам; здесь обсуждаются только две из них. Прежде всего она представляет собой ценный диагностический инструмент и фактически является составной частью наиболее развитых программных средств, разработанных для применения в Лисп-машинах. Во-вторых, люди хотят разобраться в причинах получения данного вывода интересующим их конкретным человеком, что позволяет создать определенное мнение о справедливости сделанного заключения. Отнюдь не следует ожидать, что машина полностью заменит человека-эксперта и освободит его от необходимости проверять правильность ее собственных заключений. Однако указанная

выше возможность в любом случае создает у пользователя чувство уверенности в своем компьютере.

Рассмотрев сходства и различия между символьными и цифровыми вычислениями, читатель может понять некоторые из способов, с помощью которых символьные вычисления могут быть применены к реальным практическим задачам. На основе ранее проведенного обсуждения вопросов представления знаний и стратегий поиска в следующем разделе будет показано, как эти методики приобретают конкретный вид и могут быть трансформированы, а также каким образом они могут привести к затруднениям при работе. Это особенно справедливо по отношению к процессу поиска, который очень часто является причиной возникновения «узких мест» в системах ИИ. Обнаружение путей, по которым оптика может быть применена для выполнения этих операций, может быть ключом к оптическим символьным вычислениям. До обсуждения данного вопроса рассмотрим круг проблем, связанных с использованием символьных вычислений в задачах распознавания речи, системах технического зрения и распознавания образов, обработки естественного языка и экспертных системах.

10.3. Функциональные возможности символьных вычислений

В предыдущем разделе изложены основополагающие принципы символьных вычислений, даны необходимые вводные сведения, рассмотрены способы представления данных и возможные стратегии поиска. В данном разделе обсуждение указанных проблем получает дальнейшее развитие и рассматриваются четыре главных вопроса, составляющие суть символьных вычислений: распознавание речи, техническое зрение, распознавание естественного языка, экспертные системы. Эти прикладные задачи могут принимать различные формы, каждая из которых включает сбор знаний, проведение рассуждений и поиск данных. Общим аспектом является использование знаний о системе или рассматриваемой области для улучшения понимания машинной входной информации.

10.3.1. Обзор работ по символьным вычислениям

Как и для любой новой области, вокруг проблемы ИИ возникают серьезные споры о том, что составляет суть ИИ и какой должна быть таксономия составляющих ИИ дисциплин. Усложняет ситуацию, однако, отсутствие достаточного уровня понимания общей природы задач программирования и оптимальных вычислительных структур. Тем не менее уже на данном этапе удалось добиться значительных успехов за счет применения к

широкому кругу проблем методик работы со знаниями, описанных в предыдущем разделе. В результате этого исследования символьных вычислений в настоящее время в основном сконцентрированы в четырех областях, образующих основу для создания семейства машин, обладающих способностями понимать информацию, а именно: распознавать и понимать речь, иметь техническое зрение или понимать изображение, понимать естественный язык, включать элементы экспертных систем, уметь проводить рассуждения.

Чтобы сделать машины способными идентифицировать устную речь и реагировать на нее, можно применить усовершенствованную методику распознавания образов для обработки входного сигнала и идентификации слов. Этот аспект исследований речевых систем называется распознаванием речи. Речь обычно вводится в систему распознавания речи с помощью микрофона [4]. Распознавание может осуществляться несколькими путями, но в любом из них обычно применяется выполняемая в цифровом виде процедура сравнения входной фразы или предложения с элементами, хранящимися в памяти компьютера. С другой стороны, понимание речи основывается на обработке знаний, и его относят к речи высокого уровня. Цель всех этих исследований речи, очевидно, состоит в создании высокоточной системы распознавания, полностью не зависящей от говорящего и работающей с большим словарным запасом.

Техническое зрение, более широко известное в компьютерной технике как понимание зрительных образов, относится к способности машины или компьютера понимать сцены, поступающие по визуальному входному каналу. Назначение таких систем [5] состоит в понимании образов и изображений с такой же степенью точности, как и в системах человеческого зрения. Распознавание может осуществляться большим числом способов, но в большинстве случаев используется сравнение наблюдаемой сцены с объектами, представленными в базе знаний системы. Ситуация близка к задаче создания систем понимания речи, за тем исключением, что информация заключена во входных образах, а не в форме волновых сигналов, и, кроме того, сами объекты являются трехмерными. Так как большинство входных визуализирующих устройств, например полупроводниковые телевизионные камеры, создают двумерные матрицы изображений, то для достижения определенного уровня восприятия чисто трехмерной информации, содержащейся во входной сцене, требуется прикладывать большие усилия. По аналогии со случаем обработки речи, функции обработки изображений, такие как предобработка, восстановление изображений или градиентные вычисления, называются зрением низкого уровня. Любые виды обработки, требующие взаимодействия с базой знаний, относят к зрению высокого уровня.

Понимание естественного языка основывается на способности машины определять значения фраз английского языка (или какого-то другого), вводимых в нее через периферийные устройства, обычно терминала (рис. 10.10). Следовательно, чтобы быть эффективной, система понимания естественного языка должна включать в себя знание таких разделов теории лингвистики, как анализ предложения на основе правил синтаксиса языка, синтаксический анализ и присвоение значений словам или фразам (семантики) [6]. Таким образом обработка естественного языка является процедурой, требующей очень большого числа операций со знаниями, и основанной на знании грамматики и контекста, используемых для получения значений входных предложений и фраз. Знание этих и других лингвистических свойств должно быть включено в базу знаний системы, и проблема, как и в других дисциплинах ИИ, заключается в подборе и реализации одной из многих возможных стратегий интерпретации входных сигналов (например, предложений на английском языке).

Экспертные системы являются дисциплиной, входящей в число практических приложений систем ИИ, и постепенно начинают конкурировать с возможностями людей-экспертов в специализированных областях знаний, получивших название проблемных областей. Такие системы используются для интерпретации спектральных данных (система DENDRAL) [7], разработки схем расположения элементов миникомпьютеров в зависимости от видов компонент (система R1) [8], диагностики

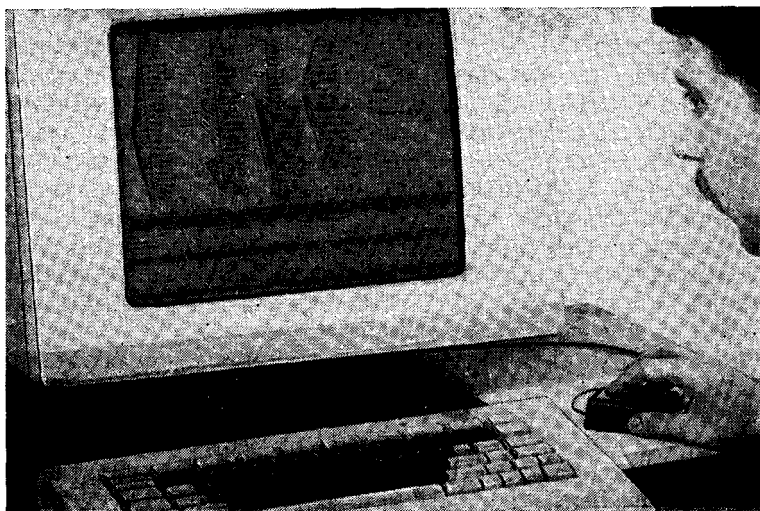


Рис. 10.10. Система понимания естественного языка (с согласия фирмы ВВН).

терапевтических заболеваний (система MYCIN) [9]. Эти компьютерные системы «достигли» высокого уровня рабочих характеристик при решении таких задач, которые требуют от человека многих лет работы, специального обучения и большого опыта [10]. В данном случае экспертиза определяется как «набор способностей, доступных высококвалифицированным специалистам-экспертам, включая наличие обширных знаний в соответствующей области эвристических правил, упрощающих и улучшающих подходы к решению задач, включая метазнания и метавосприятия, а также специальные навыки, позволяющие экономно выполнять работу» [10]. (Метазнания — знания о знаниях). Эти системы недавно были очень успешно использованы для решения задач и заданий, требующих больших объемов знаний и мощных эвристических подходов, которые обычно для своего выполнения требовали от специалистов-экспертов от 8 до 40 часов работы [10]. Используя возможности машин, работающих на языке Лисп, в сочетании со специализированными программными средствами, экспертные системы поиска информации в достаточной степени высвобождают труд экспертов, чтобы использовать его в более сложных, более нужных и разнообразных задачах.

Порядок описания функциональных возможностей, обсуждаемых в данном разделе, определяется несколькими причинами. Исторически первое успешное применение разработок в области ИИ концентрировалось на создании интерфейсов для устного или визуального сопряжения компьютера с человеком и для компьютерных игр. (Для детального обзора ранних исследований ИИ читателю следует обратиться к [6], т. 1.) Это повлекло за собой появление так называемой обработки низкого уровня, включающей в основном обработку сигналов и изображений; при этом методикам, основанным на использовании знаний, естественно отводят роль обработки высокого уровня. На основании истории развития этих дисциплин, исследования систем обработки речи и технического зрения с полным правом считают более трудными задачами, чем обработку естественного языка или создание экспертных систем. В качестве иллюстрации сказанного отметим, что многие из концепций, необходимых для обработки естественного языка, получили свое развитие в теории лингвистики и исследованиях речи высокого порядка. В свою очередь возникновение возможностей давать разъяснения в экспертных системах происходило на основе исследований систем обработки естественного языка.

Следует заметить, что происходит естественная эволюция возможностей символьных систем, которые изначально сводятся к интерфейсным устройствам человек-машина, а затем перерастают в возможность проведения рассуждений при минимальном участии человека. Данное обстоятельство позволяет

существенно упростить изложение материала, поскольку в дальнейшем будет показана большая степень общности между всеми этими дисциплинами. Во многих случаях это является результатом переноса основных идей из одной области в другую. Использование архитектуры, включающей память общего доступа, обсуждаемой в разд. 10.3.2, является именно таким случаем. Hearsay, впервые разработанная для системы понимания речи [4], была успешно применена в системах обработки естественного языка и экспертных системах.

Ниже более детально обсуждается каждая из этих способностей символных вычислений, при этом особое внимание уделено сути лежащих в основе этого методик. После этого для каждой из областей применения символных систем будут описаны основные трудности и недостатки существующих систем. Каждый из подразделов завершается кратким обзором интеллектуальных возможностей существующих и проектируемых машин, а также анализом проблем и критических параметров вычислительных систем.

10.3.2. Понимание речи

При разработке систем понимания речи основные усилия были приложены к обеспечению компьютерных систем необходимыми знаниями и разработке методов фильтрации посторонних шумов с целью достижения понимания машиной непрерывной речи в реальном времени. К большому сожалению для разработчиков компьютерных систем люди произносят слова не одинаково. Таким образом даже системы, ориентированные на узнавание речевых входных сигналов (без обязательного понимания ее), сталкиваются с задачей различать слова независимо от личности говорящего, скорости его речи, любых диалектов или акцентов, словарного запаса, пропущенных слогов и, наконец, что не менее важно, фонового шума окружающей среды.

Это, очевидно, представляет трудную задачу, поскольку сами люди не обладают надежной на 100% не зависящей от говорящего системой понимания речи. Существуют проблемы понимания речи, связанные с тем, что соседние слова произносятся слитно, что типично для произношения географических названий, таких как «Лонг Айленд», которые часто произносятся как «Лонгойлэн». Так же возникают трудности с различными диалектами, где слово «ойл» (нефть) произносится в разных частях страны как «ол» или как «оул», и очень часто приходится из контекста подбирать значения словам или пропущенным фразам. Это последнее обстоятельство включает в себе большие трудности для поиска знаний и проведения рассуждений в речевых системах, поскольку предъявляет серьезные требования к использованию методик сравнения с образцом для

идентификации слов, а также к использованию знаний для понимания значений слов и их интерпретации. Вопрос состоит в том, как и где применить знания (рис. 10.11).

Был разработан ряд методик, основанных на использовании знаний в речевых системах, но двумя наиболее совершенными из них являются системы распознавания отдельных слов (IWR) и понимания непрерывной речи (CSU) [11]. Система распознавания отдельных слов основывается на идентификации слов при раздельном произнесении слов во входном речевом сигнале, как, например,

«гелий-неоновый...лазер...сломался» или
«погода...сегодня...дождливая».

Это делается с целью уменьшения числа проблем, связанных с идентификацией начал и окончаний слов. Задача идентификации слова из преобразованного акустического входного сигнала все еще остается, и в этом случае она может быть выполнена методом сопоставления с образцом. Данный процесс (рис. 10.12) включает в себя прием входного сигнала, идентификацию основных признаков, таких как вариации основных звуков, а также начала и окончания слова. Следующий шаг — динамическое изменение временного масштаба, за счет чего подстраивают длину произнесенного слова к длине образца (шаблона), используемого для сравнения. Примером этого являются слова «гелий-неоновый» и «Карибское море», которые иногда произносятся так:

«гелий-неоновый» «Карибское море»
«ге-е-лий не-о-навый» «Ка-риб-ско-е мо-ре»
«ге-лий ни-оновый» «Ка-рибское море»,

при этом каждый из вариантов произносится с разной скоростью.

Для приведения длин слов к стандартной величине сигнал, являющийся функцией времени, расширяют или сжимают, чтобы сделать его совместимым с опорной длиной образца. Особенности этого «стандартизированного слова» затем сравниваются с образцом, хранящимся в памяти системы, а для сравнения слова с его копией в памяти используются взвешенные количественные метрики, такие как показаны на рисунке.

На данном этапе основные трудности встречаются в процессах низкого уровня. Представления о требованиях к скорости вычислений дает тот факт, что обычно на расшифровку одного слова требуется около 500 вычислений количественных коэффициентов. При этом динамическая система изменений масштаба времени приводит данное слово к стандартному фрагменту речи длительностью 500 мс [13]. Для количественных оценок можно полагать, что полное число операций умножения или

вычисления внутреннего произведения на одно слово составляет от 5 до 10 К, при условии использования взвешенной евклидовой метрики для описания признаков ряда соответствующих фильтров [14], либо применения метрики Итакуры для описания коэффициентов линейного прогнозирования соответствующих признаков [15]. Следовательно, включение оптимальной процедуры вычисления внутреннего произведения или использование оптических методов в качестве составной компоненты

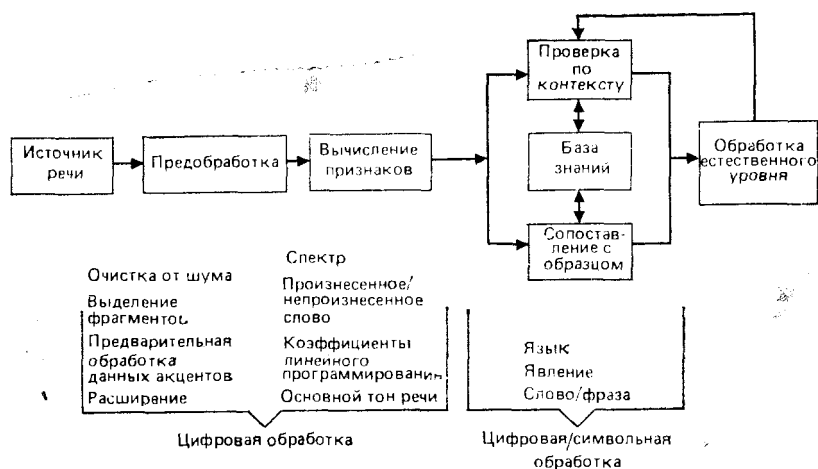
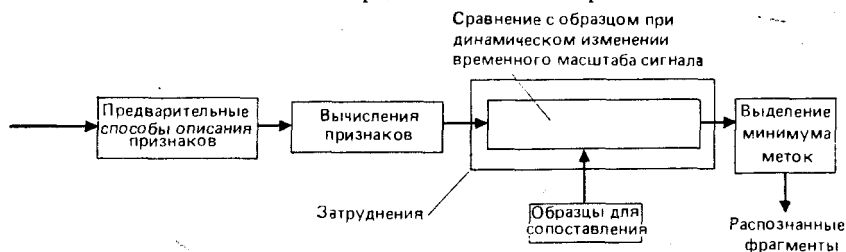


Рис. 10.11. Парадигмы понимания речи.



Метрики сравнения с образцом

$$\sum_{i=1}^{N_F} (D_i - T_i)^2 W_i$$

(Эвклидовы весовые коэффициенты для банка данных отфильтрованных признаков)

Число признаков = $N_F - 8 - 15$

$$\frac{\sum_{i=1}^{N_F} \sum_{j=1}^{N_F} D_i R_{ij} D_j}{\sum_{i=1}^{N_F} \sum_{j=1}^{N_F} T_i R_{ij} T_j}$$

(Метрики Итакуры)

Рис. 10.12. Обработка речи низкого уровня.

в речевой системе должно быть направлено прежде всего на уменьшение суммарной вычислительной нагрузки.

К сожалению, процедура сопоставления с образцом не является единственным критическим местом в процессе вычислений. Применение знаний высокого уровня, к которому относятся задачи поиска данных и проведения рассуждений, также ограничивает точность и быстродействие системы, не давая возможности эксплуатировать систему в реальном времени. На рис. 10.13 приведен пример того, как знания используются для определения значений подаваемой на вход фразы. В данном случае успешное сопоставление со словами, содержащимися в сети, позволяет машине интерпретировать фразу. Читателю следует иметь в виду то, что приведенный пример является только одним из числа возможных схем построения речевых систем. Уже появились речевые системы, использующие фреймы, сценарии и специальные правила обработки наряду с семантическими сетями.

Многие виды знаний, используемые для понимания речи, определяются исходя из лингвистики языка, и, следовательно, можно ожидать, что они будут также важны и для обработки естественного языка. Каждый из этих видов знаний характерен для используемого словаря и сильно зависит от правил грамматики языка. Таким образом, для машины важно понимать взаимосвязь прилагательного и существительного, наречия и глагола и т. д.

Наиболее очевидным и наиболее знакомым для всех типов знаний является фонетика, относящаяся к физическим характеристикам звуков в каждом слове в словаре и изучающая таким образом акустические характеристики слова. Другим важным типом знаний является морфология, рассматривающая пути, по которым основные структурные блоки, составляющие сло-

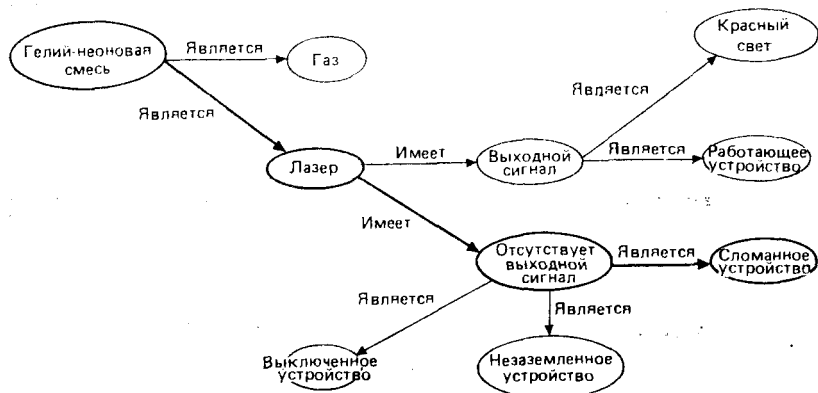


Рис. 10.13. Семантическая сеть распознавания речи.

ва (основные морфемы), могут быть скомбинированы для образования новых слов, форм множественного числа, временных форм и т. д. Используя эти типы знаний в качестве основы для организации знаний высокого уровня, можно приступить к задаче определения значения входного речевого сигнала. В начале данного раздела при обсуждении естественного языка были представлены концепции синтаксиса (структура предложения и грамматика) и семантики (способы, по которым значения слов объединяются для образования значений, предложений и фраз). Ясно, что эти виды знания необходимы для любой системы распознавания речи.

Последним и может быть наиболее важным для речи высокого уровня типом знаний является прагматика. Прагматика относится к правилам разговора и диалога, которые позволяют системе различать вопросы и сомнения в фактических ситуациях. В следующих примерах все ситуации имеют одни и те же синтаксические и семантические значения, хотя каждое из них предполагает различный тип взаимодействия с машиной:

«В гелий-неоновом лазере что-то вышло из строя».

«Что-то вышло из строя в гелий-неоновом лазере?»

«Что именно вышло из строя в гелий-неоновом лазере?»

Первая фраза представляет простое фактическое утверждение, в то время как вторая является подтверждением первой, требуя от системы четкого ответа «да» или «нет». В третьем случае, однако, ответ «да» или «нет» недостаточен; от системы требуется ответ другого уровня. Знание прагматики используемого словаря позволяет системе распознавать и понимать различия в значениях фраз и предложений.

Эти типы знаний могли быть включены в ранее приведенный пример системы распознавания отдельных слов, но этого в данном случае не требовалось. Схема использования знаний в примере независимого распознавания отдельных слов носит исключительно общий характер, применима к большинству типов знаний и называется парадигмой восходящей иерархии. В данном случае восходящая схема подразумевает использование цифровых методик, осуществляемых в начале алгоритма для улучшения качества сигнала и выделения всех признаков, что в целом составляет обработку речи низкого уровня (рис. 10.11). Последующая обработка высокого уровня состоит в использовании знаний, которые в данном случае возникают после завершения обработки низкого уровня. Понятие иерархии характеризует переход от процедур управления низкого уровня через несколько промежуточных этапов к управлению, при котором осуществляется взаимодействие знаний.

Данная парадигма не является единственной и, как для случая обработки непрерывной речи, может быть отнесена с точки зрения ее структуры к нисходящей иерархии или даже

схеме выведения среднего. В каждом из обсуждавшихся случаев, однако, остались без внимания соображения о том, где и когда знания о проблеме должны быть найдены и использованы. Если это происходит в начале процедуры, то речь идет о парадигме нисходящей иерархии; если это происходит в конце процедуры, как в случае распознавания отдельных слов, то это относится к восходящей схеме.

В системах понимания естественной речи для ограничения процесса поиска знаний используются грамматические или лингвистические модели, что означает ограничение активного словаря до определенного числа возможных слов в любой момент времени. При ограничении размера пространства поиска экономится не только время обработки для каждого слова, но и снижается вероятность неудачи и достигается рост скорости распознавания.

Основная система распознавания естественной речи состоит из акустического процессора, выполняющего преобразование сигнал-символ, и лингвистического декодера, применяющего знание для понимания вводимой речи. Типичные системы распознавания естественной речи могут быть реализованы в двух вариантах: во-первых, с распознаванием отдельных слов при случайном их следовании (как, например, для поиска в базе данных) и, во-вторых, понимание значения непрерывно произносимых предложений. В обоих случаях имеются не только общие свойства, аналогичные свойствам систем распознавания отдельных слов, но и различия. Однако главное различие заключается в разных подходах к разделению слов на фрагменты.

Разделение слов на фрагменты, или определение начала и конца отдельных слов, является критической проблемой для систем понимания непрерывной речи. Поскольку слова, произносимые в непрерывной речи, имеют тенденцию сливаться между собой, то алгоритмы обнаружения конца слова, используемые в распознавании отдельных слов, не работают. Существует несколько методов преодоления этой проблемы, и все они являются вариациями алгоритма динамического изменения временного масштаба, обсуждавшегося ранее. Последний алгоритм, хотя и является критической процедурой в вычислениях, но представляет менее трудную задачу, чем сопоставление с образцом или процесс поиска знаний.

Большинство систем понимания непрерывной речи построены по нисходящей схеме, ориентированной на знания и использующей знания о проблеме, обсуждаемой в произносимой речи, для ограничения ожидаемого объема информации, содержащейся во входном сигнале [11]. Такая методика была успешно реализована в системе Hearsay-III [12], где каждый вид знаний может взаимодействовать с частично обработанным входным

сигналом независимо от других типов знаний, как показано на рис. 10.14. Особенностью архитектуры данной системы является наличие рабочей области общего доступа, предназначенной для связи между различными «экспертами» по знаниям. Фактически, позволяя каждому типу знаний оказывать влияние на входной сигнал, получают по существу развитую «экспертную систему» индивидуальной речи (см. разд. 10.3.5). Эта концепция области общего доступа, допускающая взаимодействие между собой большого числа баз знаний, была успешно использована как в системах понимания естественного языка, так и в экспертных системах.

Использование эвристического поиска для ограничения объема информации во входном сигнале имеет свою цену и требует установления компромисса между общностью выполняемых задач и получаемыми характеристиками. Этот вопрос составляет предмет рассмотрения данного раздела и затрагивает в настоящее время практически все вопросы исследований ИИ. Чтобы оценить трудности достижения указанного компромисса, читателю следует принять к сведению, что словарный запас английского языка очень велик — 24 тома Оксфордского словаря английского языка составляют его «священное писание». Хотя

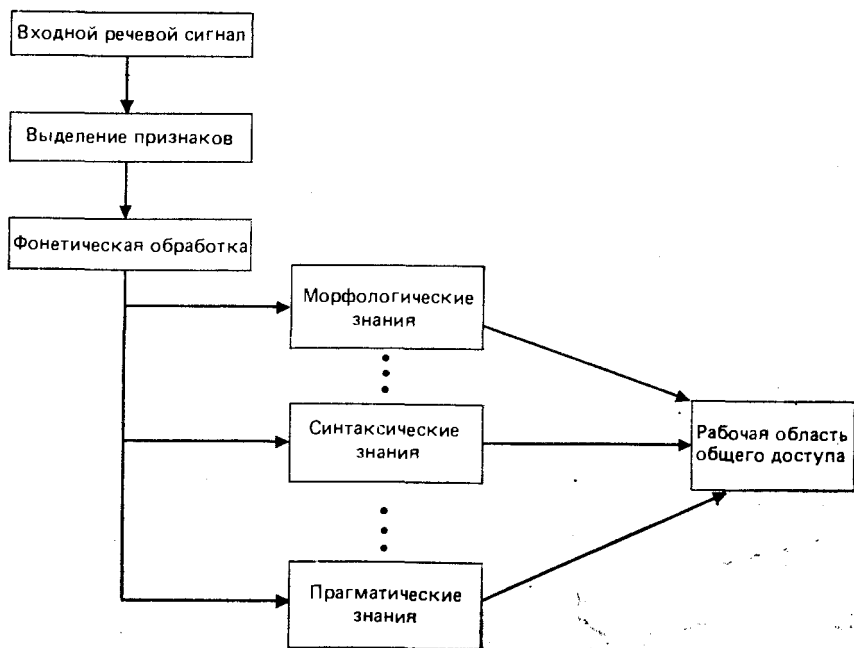


Рис. 10.14. Архитектура системы понимания непрерывной речи Hearsay-III.

большинство людей в обиходе использует намного меньший словарный запас, но даже 20 000 слов создают ошеломляющее число возможных комбинаций слов и число вариантов их произношения. Ранние речевые системы, в которые старались включить большинство слов из среднего словарного запаса человека, никогда не были способны даже приблизиться к режиму работы в реальном времени и расходовали слишком много времени на поиск данных в памяти. Решение данной проблемы, использующее аналогии с человеческим мышлением, состоит в использовании для ограничения поиска знаний, граничащих с интересующей проблемой или областью. После того как эта схема была реализована, удалось создать речевые системы, работающие почти в реальном времени, а также добиться успехов в реализации систем технического зрения, систем естественного языка и экспертных систем.

Читателю следует обратить внимание на то, что большая часть недавних усовершенствований в системах распознавания отдельных слов и непрерывной речи обусловлена успехами в создании процессоров обработки сигналов специального назначения. Такие процессоры включают в себя ИС специального назначения и систолические структуры, служащие для обработки речи низкого уровня. Остающиеся при этом критические процедуры, а именно поиск знаний и их обработка, требуют дополнительных исследований символьных вычислений и структуры оптимальной программной среды высокого уровня. (Несколько таких структур описаны в разд. 10.4, где более подробно рассматривается соотношение между структурой обработки и выполняемой функцией.) Как будет показано в следующей части раздела, процессоры специального назначения используются также в системах технического зрения.

10.3.3. Техническое зрение

Техническое зрение, как подразумевает само это понятие, заключается в создании машины, способной понимать в реальном времени визуальный входной сигнал. Как и для человеческого зрения, машинное зрение включает в себя идентификацию той информации, которая содержится в образе или визуальной сцене и которая указывает, как различные элементы связаны друг с другом в пространстве и времени. Область применения таких систем отнюдь не ограничена использованием в качестве входных устройств, облегчающих организацию связи человека с машиной, а включает в себя широкий круг проблем, как промышленных, так и военных. Последние успехи, достигнутые в автоматизации процессов и робототехнике, например создание роботов, осуществляющих выборку из бункера, стали возможными благодаря исследованиям машинного зрения. Другая об-

ласть промышленного применения подобных систем включает в себя создание сенсоров для автоматической сварки, обработку опасных для здоровья материалов, изготовление микросхем со сверхвысокой степенью интеграции, автоматизированное проектирование и производство, проверку готовых изделий, получение изображений в медицине, обнаружение удаленных объектов в картографии, управление транспортом, создание вспомогательных средств для водителей транспортных средств и авиации, разведку месторождений нефти и полезных ископаемых. Область возможных применений в военной технике может быть столь же широкой и включает в себя, например, автономные системы навигации движущихся объектов, системы обработки фотоснимков, средства ведения разведки, системы выбора цели и дальномерные устройства, средства наведения на цель, определения траекторий движущихся объектов. Все это составляет те задачи, где наиболее целесообразно применение оптических методов, поскольку визуальная информация по своей природе является оптической и может обрабатываться двумерными способами.

Во многих случаях системы машинного зрения создаются на базе систем, основанных на знаниях и описанных в предыдущем разделе при изложении вопросов создания систем понимания речи. Как и в случае речевых систем, знания высокого уровня о рассматриваемой сцене или изображении эффективно используются для ограничения объемов информации, используемой в процессах идентификации и понимания визуальных сцен. Аналогично в системах технического зрения знания высокого уровня, характеризующие сцены, могут эффективно использоваться для управления операциями низкого уровня. Ниже в данном разделе будут приведены примеры того, как знания могут применяться для ограничения процессов поиска и обработки данных.

Другая параллель между речевыми системами и системами технического зрения заключается в близости подходов к использованию знаний и управлению. В одной из ранних работ по проблеме технического зрения парадигма восходящей иерархии использовалась аналогично тому, как это делалось в случае распознавания отдельных слов. Согласно этому подходу, на стадии предварительной обработки осуществляется обнаружение края объекта, производится выделение признаков и определяется их взаимосвязь, но не используется получение логических выводов на основе имеющихся знаний. Как и в случае речевых систем, система, выполняющая этап предварительной обработки, называется системой технического зрения низкого уровня (выполняющей соответственно обработку низкого уровня). Система, включающая в себя последовательное взаимодействие с базой знаний, называется системой техниче-

ского зрения высокого уровня (соответственно речь идет об обработке высокого уровня).

В настоящее время разработчики изучают возможности сочетания обработки высокого и низкого уровня в парадигмах большего масштаба. Например, выполнение рассуждений высокого уровня об ожидаемых свойствах объекта и их взаимосвязи оказывается полезным для организации оптимального прохождения задач и управления обработкой низкого уровня. Преимущества такой схемы обработки данных определяются эффективной концентрацией усилий на конкретных участках изображения, таких как выделение края объекта или нахождение линейных элементов изображения. На других стадиях процесса технического зрения обработка низкого уровня может быть использована для выработки гипотез, которые следует проверять с помощью обработки высокого уровня. В этом случае эффективность процесса удастся повысить путем отбрасывания гипотез, которые несовместимы с имеющимся изображением.

Аналогично тому как системы обработки речи низкого уровня включают в себя методики распознавания образов, системы технического зрения также развивались на основе ранних работ по распознаванию образов. Однако рассмотрение технического зрения только как процесса распознавания образов является неточным; распознавание образов является по своей сути исключительно цифровой процедурой, выполняемой непосредственно с изображением. Техническое зрение, напротив, использует знание о сцене для того, чтобы оценить категорию изображения, и работает с символьным представлением изображения, а не с самим изображением. Это составляет наиболее существенное различие между ними, и в целом ситуация оказывается аналогичной случаю использования знаний синтаксиса для понимания языка (вопрос понимания естественно-го языка излагается в разд. 10.3.4).

После общего описания концепции машинного зрения авторам хотелось бы дать элементарное представление о типах процессов и вычислений, используемых в системах технического зрения. Детальное обсуждение выходит за рамки данной главы; за дополнительной информацией по этому вопросу читателю следует обратиться к таким прекрасным книгам, как [5] и [13].

Для простоты в последующем обсуждении проблем технического зрения обработка будет рассматриваться как процесс, происходящий в элементах детектора изображения, который обычно представляет собой тот или иной вид двумерного матричного устройства, например телевизионной камеры или видикона. Как и в случае речевых систем, техническое зрение низкого уровня в основном сосредоточивается на преобразова-

нии уровня сигнала в символьную информацию. Эта информация «записана» в виде вариаций интенсивности элементов двумерной матрицы и относительного расположения этих вариаций. Кроме того, информация может содержаться в цвете и текстуре изображения, так как оба этих признака могут быть выделены из входного изображения. Процедура обработки начинается с выделения признаков, и выполняется посредством обработки данных, записанных в ячейках, и направлена на обнаружение краев отдельных фрагментов, как это показано на рис. 10.15.

Первый шаг этой процедуры — стадия предобработки, которая подготавливает изображение для обработки. Например, может потребоваться специальная обработка изображения, связанная с наличием шумов или нерезкостью изображения, определяемых оптической схемой или механическим перемещением отдельных узлов схемы. Коррекция геометрических искажений необходима, чтобы исключить влияние угла зрения на изображение и скорректировать искажения, вносимые в изображение линзами или какими-либо другими источниками при вводе изображения в компьютер. Можно применить несколько сенсоров, чтобы получить, например, трехмерное изображение или использовать передачу информации сразу на нескольких длинах волн. В таком случае входные сигналы, поступающие с нескольких сенсоров, могут быть объединены на стадии предобработки с целью достижения максимально надежной работы системы.

Обычно первой операцией после предобработки является операция выделения признаков. Эта процедура, как правило, выполняется путем нахождения градиентов в изображении; для этого вычисляют разность двух гауссовых распределений. Данная процедура иллюстрируется на рис. 10.16. Она реализуется путем вычитания двух гауссовых функций, в результате находят две точки перехода через нуль. Указанная процедура позволяет получить распределение градиентов интенсивности света и, в частности, идентифицировать края объекта.

Следующий шаг обработки называют выделением общего контура объекта и фрагментацией изображения первого порядка. Выяснив, где находятся края объектов, имеющих в сцене, какие линии и края объектов в одной части сцены являются продолжением линий и изображений в других частях сцены, какие области сцены принадлежат отдельным объектам, с помощью выделенных признаков удастся определить пространственный контур объекта. Идея здесь заключается в как можно более быстром переходе от вычислений данных, содержащихся в отдельных ячейках, к рассмотрению целого объекта.

В качестве примера на рис. 10.17, а—г представлен танк, движущийся по дороге. Оператор вычисления разности гауссо-

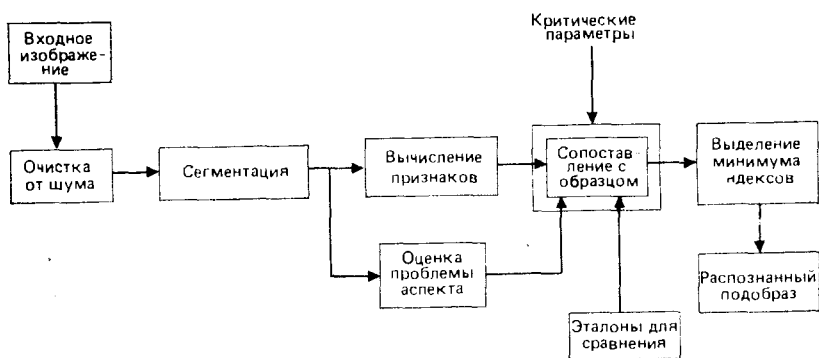


Рис. 10.15. Обработка низкого уровня в системе технического зрения (сопоставление с образцом).

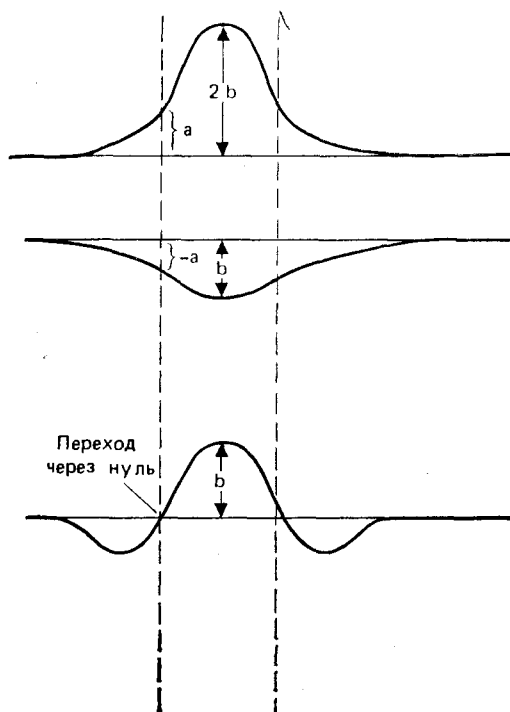


Рис. 10.16. Операторы вычисления разности гауссовых распределений.

вых распределений выделяет края изображения танка, которые на этапе выделения общего контура объекта могут быть соединены линиями. В примере с танком его изображение делилось на отдельные области, содержащие башню, корпус, дорогу, гусеницы.

Необходимы специальные средства, позволяющие распознать конкретные объекты по их общему контуру (для танка — это гусеницы, пушка, дорога и т. д.), а также устанавливать соотношения между этими объектами. Это делается в процессе выделения признаков изображения и на стадии их разбиения по группам. В настоящее время имеется несколько конкурирующих между собой способов выполнения этой процедуры, которая хотя и просто описана на словах, но требует большого объема сложных цифровых и символьных вычислений. Символьное представление разбитого на фрагменты изображения

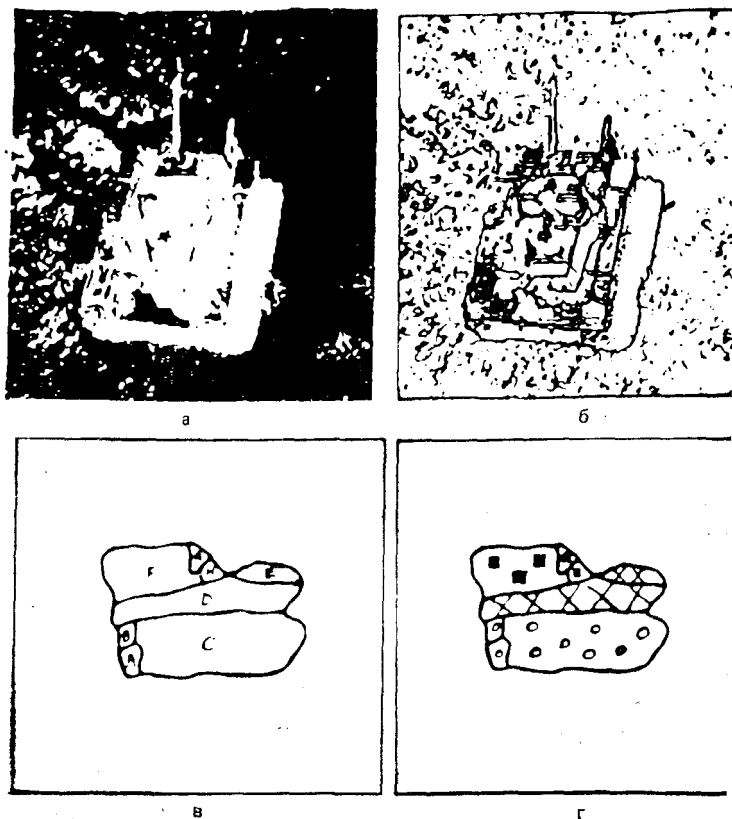


Рис. 10.17. Пример обработки изображения на промежуточных этапах в системе технического зрения.

осуществляется на следующем этапе обработки, и в конце концов получающийся набор символов краев и различных областей получает семантическую интерпретацию, что дополнительно требует применения ряда моделей, создаваемых на самом высоком уровне обработки. Общая архитектура рассмотренной системы представлена графически на рис. 10.18.

В качестве примера описанного выше процесса можно рассмотреть задачу идентификации того факта, что изображение на самом деле является танком. Один из аспектов этого процесса заключается в определении идентичности и соотношения между собой фрагментов изображения на рис. 10.17. Методика выполнения этой задачи, основанная на использовании фреймов (разд. 10.2.2), представлена на рис. 10.19. Здесь для выяснения формы башни и ее положения относительно корпуса проводится анализ подобных объектов, содержащихся в базе знаний. В рамках фрейма башня рассматривается как производный объект от класса объектов типа пушка, имеющий два конкретных воплощения: башня танка и башня корабельного орудия. Этот фрейм может вызывать образ башни из памяти системы с целью символьного сравнения с фрагментированным исходным изображением. Тогда путем рекурсивного повторения этой процедуры система может идентифицировать корпус танка. Система может построить гипотезу, предполагающую, что фрагменты изображения являются элементами танка. Но данная гипотеза не будет подтверждена до того момента, как будут определены другие элементы сцены.

Как можно ожидать, данный процесс представляет собой весьма серьезное «узкое место» в системах машинного зрения даже в случае применения знаний в соответствующей области для ограничения пространства поиска. Дело в том, что получение единственного непротиворечивого вывода может потребовать использования большого числа гипотез и операций сравнения. Это наводит на мысль о целесообразности использования для таких вычислений параллельной обработки, но, к сожалению, в настоящее время в реальном времени такие процедуры проводить не удастся. Однако, если объект идентифицирован, то оставшуюся часть вычислений выполняют в символьном виде.

Последнее обстоятельство приводит к такому этапу вычисления, как понимание содержания наблюдаемой сцены, которая, к сожалению, представляет наименее понятную область всей процедуры. Совершенно недостаточно указать, что объект имеет башню, пушку, гусеницы, поскольку такой объект может быть не танком, а гусеничным бронетранспортером или каким-то другим транспортным средством. Указать, что объект представляет собой танк — это опять-таки слишком далеко от понимания, о каком типе танка идет речь или какие действия он

совершает. Чтобы выполнить даже простейшее из этих заданий, потребуется применение усложненных методов моделирования, связанного с наличием четкого описания этих объектов в базе знаний системы. Широко распространенные методики используют ряд изображений для анализа движения, а затем делают заключение относительно действий объекта. Однако временные характеристики параметров объекта не всегда являются доступными, следовательно, приходится разрабатывать методики понимания сцен из контекста.

В процессах технического зрения знание о сцене может быть использовано для ограничения поиска на любом из уровней

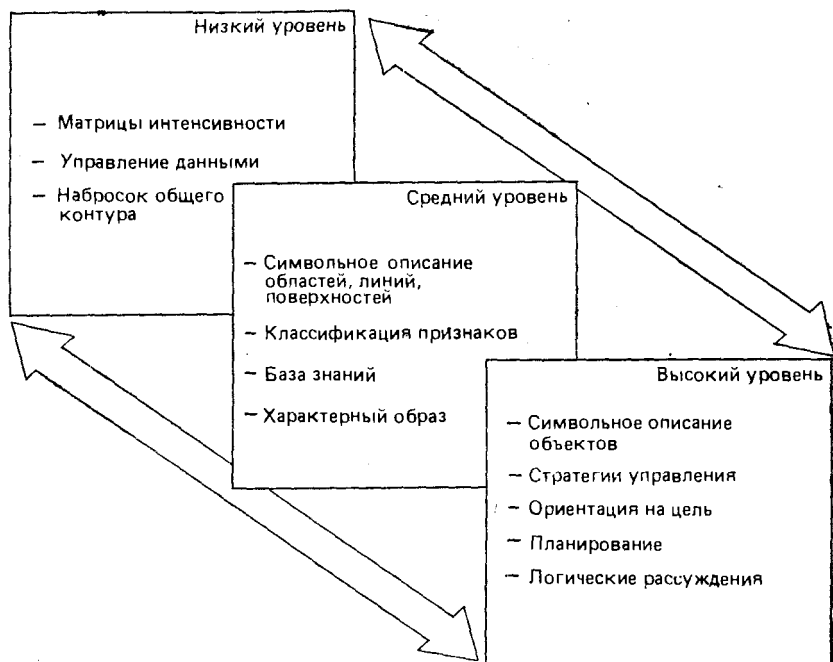


Рис. 10.18. Парадигма обработки изображения.

Фрейм:	Башня
От:	Пушка
Класс 1:	Танк
Класс 2:	Судно, вооруж. пушками
Форма:	Овальная
Слот 1:	Пушка
Следующ.к 1:	Танк. База
Следующ.к 2:	Пушечное вооружение. База
Изображение:	Модель. Башня



Изображение. Танк

Рис. 10.19. Прототип фрейма, описывающего танк,

обработки. Вначале делают заключения относительно геометрических свойств объектов на основе знаний о невидимых частях трехмерных геометрических объектов, таких, как башня танка; с этой целью совершают переход к двумерной модели. При этом знание конкретной области позволит, например, отличить шоссе от взлетно-посадочной полосы; для этого можно использовать сведения о возможных различиях окружающей среды возле шоссе и взлетно-посадочной полосы. Наконец, идентификация объекта шоссе увеличит уверенность в том, что сцена представляла собой танк, а не корабельное орудие.

Если бы использовалась какая-либо иная стратегия управления, отличающаяся от восходящей, то имеющиеся знания о сцене ограничивали бы действие системы иным образом. Как и в случае обработки речи, возможны самые различные виды стратегии управления процессом понимания изображения, включая нисходящую иерархию, смешанные подходы нисходящей и восходящей иерархий и различные варианты подхода, использующего рабочую область общего доступа. Для нисходящей иерархии контролируются и проверяются предсказания, сделанные с помощью моделей высокого уровня, содержащихся в базе знаний (например, предположение о наличии в изображении такого объекта, как дорога). В этом случае для проверки гипотез высокого уровня обычно используются специальные образцы для сопоставления. В случае организации в системе рабочей области общего доступа процессоры, выделяющие признаки, и процессоры, выполняющие символьные и семантические процедуры, работают параллельно и взаимодействуют друг с другом посредством общей для всех них рабочей области памяти. Важным моментом здесь является то, что имеющиеся знания влияют на все уровни вычислений, какая бы стратегия управления ни использовалась, и очень часто для проверки любой гипотезы на низших уровнях обработки требуется применение итерационных процедур.

Из сказанного выше читателю следует понять, что основу проблемы технического зрения составляет распознавание сложных объектов. Для устранения имеющихся в настоящее время «узких мест» требуется развитие новых более эффективных алгоритмов и гибких методик сопряжения различных этапов обработки, таких, как обнаружение основных компонент изображения, выделение/группирование признаков, взаимодействие с базой знаний. Проблема состоит в том, что в настоящее время большинство алгоритмов технического зрения работает очень медленно и отдельные стадии обработки разобщены и очень часто требуют применения различных аппаратных средств для достижения желаемых функциональных средств. С этим связана необходимость эффективного представления знаний в памяти системы в форме, доступной не только для

определенных вариантов операции сопоставления с образцом, но также для присваивания символов и определения контекста. Эффективная архитектура для систем технического зрения и обработки изображений, включающих понимание объектов, должна соответствовать всем этим требованиям. Как представляется авторам, за счет использования параллелизма в вычислительных операциях систем технического зрения и разработки соответствующих систем, реализующих алгоритмы технического зрения, многие из этих трудностей могут быть преодолены. Оптические способы решения этой задачи, как будет показано в разд. 10.4, также являются перспективными.

Авторы уверены в том, что системы технического зрения, использующие методики моделирования высокого уровня, в случае применения оптических методик могут получить ряд преимуществ, поскольку они объединяют элементы как цифровых (например, обнаружение/усиление края объекта, фурье-методики для вычислений признаков), так и символьных (сопоставление с образцом, распознавание объектов и т. д.) вычислений [14]. Так как такие системы требуют интенсивной работы с базой знаний, то применение систем с большой пропускной способностью памяти (см. разд. 10.4) может облегчить решение некоторых задач, связанных с итерационным процессом идентификации. Наиболее совершенные системы понимания изображений [15, 16] требуют сопряжения чисто цифровой обработки изображения, выполняемой на самом низком уровне выделения фрагментов изображения, с символьными вычислениями, выполняемых на более высоких уровнях в процессе классификации и распознавания объектов [17]. Сочетание цифровых и символьных вычислений для решения таких задач технического зрения, как управление оптическими потоками, может явиться для оптических вычислений наиболее важным применением в области ИИ.

В данном разделе представлено введение в проблему машинного зрения и на ряде примеров выявлены некоторые из наиболее трудных вычислительных проблем, связанных с процессами технического зрения. В то же время проведены параллели между обработкой изображений и обработкой речи, используя в качестве основы для обсуждения проблему понимания речи, рассмотренную в предыдущем разделе. В следующем разделе будут представлены новые аспекты ряда уже обсуждавшихся вопросов, таких, как иерархия управления и проверка гипотез.

10.3.4. Понимание естественного языка

Часто задают вопрос: «Что такое обработка естественного языка и чем она отличается от понимания речи?». Чтобы ответить на него, сначала определим понятие понимания естественного

языка (ПЕЯ), а затем исследуем его взаимосвязь с различными типами знаний, используемых в понимании речи. После этого, с помощью примера, взятого из хорошо известного литературного источника, будут рассмотрены некоторые из методик, используемых в системах ПЕЯ. На этой основе в разд. 10.3.5 будет описан метод экспертных систем.

Наиболее известное определение естественного языка, являющееся столь же точным, как и все другие определения, состоит в том, что это язык, используемый в устной или письменной форме в качестве основного средства связи между людьми. Отсюда в ряде стран и в большей части США естественным языком обычно называют английский, в других странах это испанский, китайский или какой-либо еще. Из этого следует, что лингвистические признаки играют важную роль в обработке знаний, связанных с естественным языком, и что естественный язык тесно связан с областью вычислительной лингвистики.

Как это уже имело место для систем обработки речи и технического зрения, ПЕЯ имеет свою историю развития в области исследования интерфейсных устройств для сопряжения человека с машиной. Здесь задача состояла в создании машины, понимающей фразы или предложения на английском языке, вводимых в нее с помощью каких-либо периферийных устройств (обычно клавиатуры). На ранних этапах развития этого направления основной целью (помимо перевода языка) являлось создание систем, позволяющих обращаться с запросами к большим базам данных. Сама структура запроса и его интерпретация имели очень большое значение, и пользователю следовало быть очень внимательным, следя за точным переводом запроса на язык базы данных. Для приведенного ранее в данной главе примера запрос по обнаружению в базе данных сведений, посвященных нелинейным оптическим материалам для двумерных пространственных модуляторов света, мог бы, в случае ввода сигнала на естественном языке, иметь следующую структуру:

«Найти все сообщения по нелинейным оптическим материалам, применяемым в двумерных пространственных модуляторах света».

Первые варианты таких систем были способны интерпретировать такую входную команду только в том случае, если все эти слова входили в словарный запас системы. При соответствующей системе индексирования эта входная команда трансформировалась бы так:

Класс: двумерный (w) пространственный (w) модуляторы света (w)

Подуказатель: нелинейный (w) оптический (w) материал где (w) относится к связкам слов. Даже в настоящее время

многие системы запросов к базам данных все еще действуют таким же образом. Чтобы далее проиллюстрировать это утверждение, предположим, что требуется получить сообщения о рабочих характеристиках двумерных ПМС. Запрос

«Требуется найти все сообщения об их рабочих параметрах» привел бы только к появлению сообщения об ошибке, даже если бы данное утверждение непосредственно следовало за первым. Это обусловлено тем, что вторая фраза содержит косвенное указание на элемент первого предложения, так что подлежащее второй фразы не было бы однозначно идентифицировано. В этом случае система понимания естественного языка, называемая системой диалогового языкового общения, требует использования знания прагматики (см. разд. 10.3.2), что позволяет облегчить решение указанных выше проблем.

Таким образом, при обсуждении ПЕЯ речь по существу идет о реализации возможностей, аналогичных таким человеческим способностям, как умение читать и понимать тексты, вести диалоги. Однако не существует прямых аналогий между обработкой естественных языков и обработкой сигнала и изображения, так же как и в случае обработки речи низкого уровня и систем технического зрения. На практике используют преобразование сигнал-символ, но это применяется непосредственно на стадии ввода сигнала, а после этого момента вся обработка носит символьный характер. Однако спасает ситуацию то обстоятельство, что большая часть знаний, используемая в системе ПЕЯ, непосредственно применима для понимания речи высокого уровня.

Чтобы получить лучшее представление о способах обработки естественного языка, следует проанализировать возможности практического применения ПЕЯ. Поскольку здесь будут рассматриваться лишь наиболее важные применения, то за более полной информацией читателю следует обратиться к монографии [18]. За исключением случая использования ПЕЯ для создания входных интерфейсов, главное применение ПЕЯ лежит в области компьютерного программирования. Здесь цель состоит в замене выражений, таких, как

Do 100 I = 1,50

J = J + DATA (I)

100 Continue

AVE = J/50

на следующее выражение:

«Вычислить среднее значение для 50 величин».

В настоящее время разрабатываются языки, являющиеся более похожими на английский, особенно для задач в области ИИ;

это будет показано в разд. 10.3.5 на примере обсуждаемых в нем экспертных систем.

Другой аспект ПЕЯ состоит в обработке текстов, под которой подразумевается обработка ряда предложений или параграфов с целью извлечения важной информации. Это целенаправленное выделение информации очень полезно при работе с литературными источниками, как, например, сбор информации по оптическим вычислениям из всех журналов, издаваемых Институтом инженеров по электротехнике и радиоэлектронике (IEEE) и Американским институтом физики (AIP). Другим приложением обработки текстов служит механический перевод текстов, например с английского языка на испанский. Наконец, задача создания систем, выдающих информацию на естественном языке, так же как и реализация способностей давать разъяснения в экспертных системах, является наиболее важной прикладной задачей ПЕЯ. Сведение процесса получения заключений по какому-то вопросу к серии простых предложений представляет, однако, труднейшую проблему даже для людей.

Любая система обработки естественного языка включает ряд основных компонентов. Входной сигнал вводится с терминала, хотя иногда в этих целях используется микрофон. В этом случае, однако, микрофон применяется для ввода букв, идентифицируемых с помощью программных средств, и он не позволяет выделять отдельные слова и предложения.

Вслед за вводом входного сигнала система ПЕЯ с целью идентификации взаимосвязей и взаимозависимостей между словами разлагает или подвергает предложение синтаксическому анализу. Программа синтаксического анализа основывается на знаниях грамматики языка и направлена на обнаружение подлежащего и связей имен существительных и глаголов с производными от них формами. Например, хорошо известная фраза, заимствованная из журнала *Applied Optics* «Оптика подразумевает работу света» (*Optics is light work*), может иметь различные варианты синтаксического разбора, примеры которых показаны на рис. 10.20.

При синтаксическом разборе осуществляется разложение введенной в систему фразы и выделяются взаимосвязи между словами, которые заносятся в память в символьном виде. После синтаксического разбора семантический интерпретатор получает соответствующую информацию и, используя уже имеющуюся информацию из базы знаний, присваивает каждому слову фразы некоторый коэффициент. Это осуществляется либо в процессе поиска, либо с помощью преобразования в промежуточный формат, известный как «язык представления значений». Данный язык не изменяет значения и выраженные в виде символов взаимоотношения слов, но он реализован так, что обеспечивает более эффективное отображение на словарный запас

системы. Из приведенного выше примера следует, что система может определить истинность любого из ниже следующих высказываний:

Оптика = Не_Работа (Существительное)

Оптика = Простая (Прилагательное) + Работа (Существительное)

Оптика = Работа (Существительное) со Свет (Существительное)

Оценить каждую из этих интерпретаций можно с помощью обработки на более высоком уровне. Для этого словесная конструкция передается в процессор проблемной области и языкового общения, который, используя прагматические знания, вырабатывает гипотезу о значении введенного сообщения. Гипотеза либо сравнивается с информацией, хранящейся в базе знаний, либо требует проведения дополнительной семантической и контекстуальной обработки, а также изменения проблемной области.

В приведенном выше примере парадигма восходящей схемы решения использовалась для объяснения процесса обработки знаний аналогично тому, как это делалось при анализе систем понимания речи и технического зрения. Метод восходящей

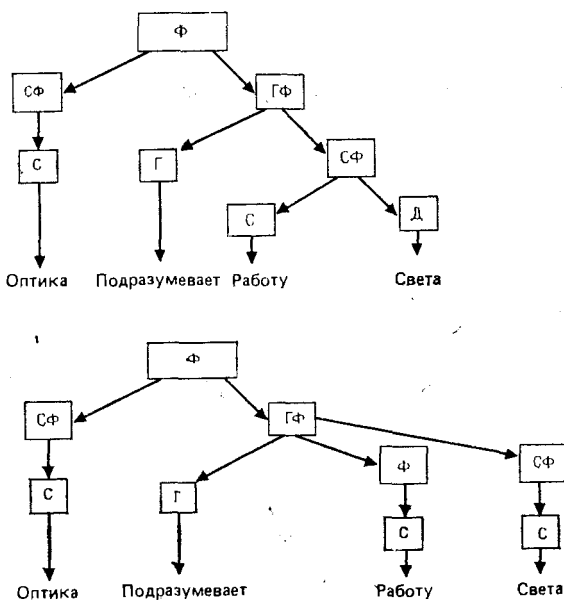


Рис. 10.20. Возможные варианты синтаксического анализа фразы «Оптика подразумевает работу света» («Optics is light work»).

Ф — фраза; СФ — фраза с существительным; ГФ — глагольная фраза; С — существительное; Г — глагол; Д — дополнение.

схемы решения показан в виде схемы на рис. 10.21. Однако возможны и другие подходы. В качестве примера можно рассмотреть архитектуру системы, использующей рабочую область общего доступа, приведенную на рис. 10.22. Она может быть использована для объединения знаний, получаемых от различных процессоров в случае неполного понимания этих знаний.

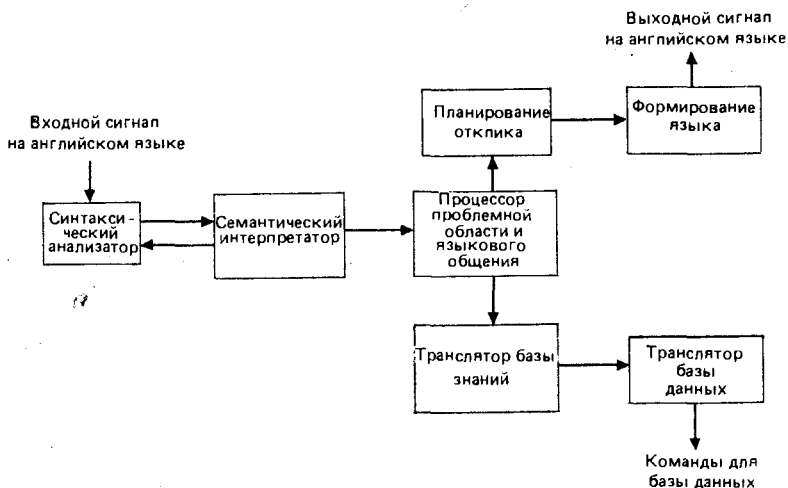


Рис. 10.21. Компоненты системы обработки естественного языка.

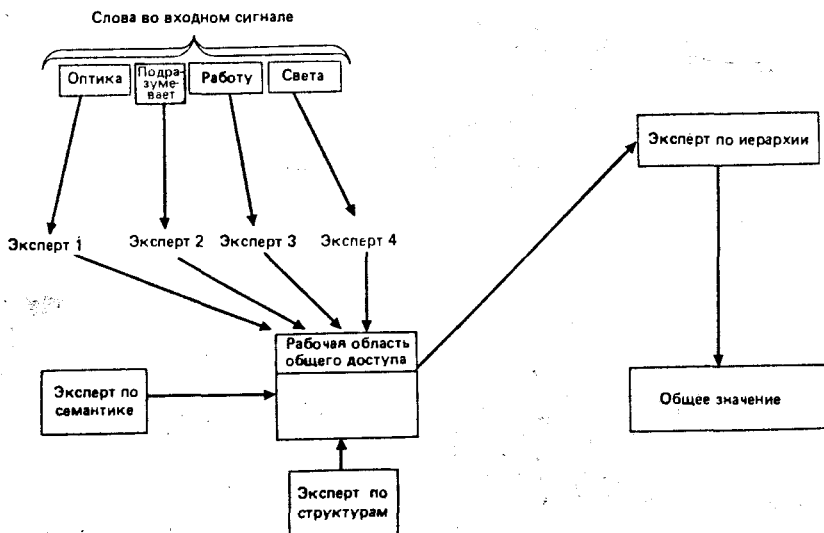


Рис. 10.22. Система обработки естественного языка с рабочей областью общего доступа.

Указанная схема позволяет быстро достичь полного понимания входных данных, реализуя это параллельными методами. Например, если синтаксический анализатор находит два различных синтаксических значения одного и того же предложения «Оптика подразумевает работу света», то используемые процессоры должны будут определить наиболее вероятную интерпретацию путем использования дополнительных знаний.

К сожалению, существующие в настоящее время системы обработки естественного языка неспособны однозначно определить значения входных предложений или эффективно использовать контекст и прагматические знания. Каждая фраза обработки имеет свои «узкие места», особенно когда вводимая информация содержит ошибки в пунктуации и орфографии. При вводе в систему обработки естественного языка неотредактированных текстов требуется предусмотреть дополнительные программные средства. В настоящее время такие системы работают медленно и их возможности ограничены пониманием текстов в очень узких областях, с низкой степенью точности интерпретации. Как это было в случае речевых систем и систем технического зрения, в системах обработки естественного языка также приходится искать компромисс между общностью поставленных задач и получаемыми рабочими характеристиками.

Общая цель исследования проблем понимания естественного языка заключается в разработке достаточно гибкой системы, обладающей перестраиваемой проблемной областью, которая может достигать высокого уровня точности интерпретации. Как это уже встречалось ранее, стремление оптимизировать рабочие характеристики положило начало разработке параллельных алгоритмов, направленных на достижение оптических или мультипроцессорных реализаций. Однако, как и в случае технического зрения, достигнутого уровня понимания принципов параллельной организации прохождения задач в настоящее время недостаточно для их реализации.

После предшествующей дискуссии можно начать рассмотрение вопроса о том, как мультипроцессор или оптический процессор может быть использован для понимания естественного языка в реальном времени. Процесс синтаксического анализа может осуществляться параллельно, так как во многих случаях различные морфемы, лексические нюансы и синтаксические структуры анализируются параллельно. Организация рабочей области общего доступа для основных процессоров незамедлительно позволила бы реализовать параллельную обработку.

Последнее, на чем следовало бы остановиться особо, заключается в том, что при описании обработки естественного языка обсуждается поведение компьютерных систем. В данном случае анализируются возможности таких систем давать интерпретацию и делать заключения относительно входных данных,

а также, если это удастся, учиться на своих ошибках. В свою очередь вопрос об интерпретации данных требует выяснения того, где заканчивается сфера обработки естественного языка и начинается обработка, присущая экспертным системам. В обоих случаях становится все более общей форма представления знаний, поскольку накопленные знания будут использоваться для понимания новых естественных языков, вводимых в систему. Общие знания об окружающем мире, т. е. общеизвестные истины, так же как и специальные значения по конкретной области, должны использоваться для определения того, следует ли вновь введенную информацию понимать буквально, или она содержит элементы метафоры, юмора или сарказма. Указанные вопросы будут затрагиваться еще раз в следующей части раздела при рассмотрении экспертных систем.

10.3.5. Экспертные системы

Последним из рассматриваемых в данной главе практических применений символьных вычислений являются экспертные системы. Работа таких систем почти полностью основывается на использовании различных процедур обработки символьной информации, что находится в резком противоречии с ранее обсуждавшимися системами. В системах обработки речи, технического зрения и ПЕЯ упор делался на определенные формы преобразования сигнала в символ, что связано с использованием знаний высокого уровня. Во многих случаях применяются процессы поиска знаний и рассуждения. Экспертные системы в дополнение к поиску знаний и проведению рассуждений используют также процесс сбора новых знаний. Несомненно, что успехи в других сферах принесли бы значительную пользу для развития экспертных систем, развивающихся на стыке этих смежных дисциплин.

Но что такое экспертная система? В наиболее широком смысле экспертная система представляет собой машину, которая имитирует или подражает процессам человеческого мышления или умению проводить рассуждения. Такая система старается использовать методики принятия решений человека-эксперта для решения конкретных проблем. Этими методиками во многих случаях являются эмпирические или эвристические правила, описанные в разд. 10.2. Метод анализа задачи экспертом, рассматриваемая им информация, необходимые данные, знания о знаниях, которыми он располагает (то, что называется метазнания), информация, отбрасываемая в процессе поиска, — все это является элементами экспертизы. В экспертных системах стараются реализовать всю совокупность функций и применить их для решения конкретной задачи. Цена, которую придется заплатить за это, как было показано в предыдущих

разделах, состоит в потере универсальности системы; до настоящего времени экспертные системы были способны работать лишь в узкоспециализированных областях. Некоторые из этих успешно осуществленных вариантов ранее уже упоминались в данном разделе, а именно: системы, сконструированные на мини-компьютерах серии VAX и PDP [18], системы интерпретации спектроскопических данных DENDRAL [7], система предназначенная для проведения терапевтической диагностики MYCIN [9].

В каждой из этих систем знания загружаются в машину, которая способна понять задачу, в значительной степени аналогично тому, как это происходит в системах обработки речи, технического зрения и естественного языка. В этом случае машина выполняет функцию принятия решения. Для выполнения функций эксперта система использует поиск знаний и механизм рассуждений, получая при этом заключения по проблеме на основе данных, предоставляемых пользователем. Мощность этих экспертных систем определяется объемом символьной информации, которую они могут накапливать в базе знаний, и скоростью ее обработки.

В любой экспертной системе имеются три компонента (рис. 10.23): база знаний, механизм логического вывода и средства обоснования. В процессе экспертизы данные накапливаются в базе знаний, при этом используется один или более типов представления, описанных в части разд. 10.2.2. База знаний может включать в себя элементы эвристики и метазнания. Механизм логического вывода дает экспертной системе возможность проводить рассуждения, позволяя системе группиро-

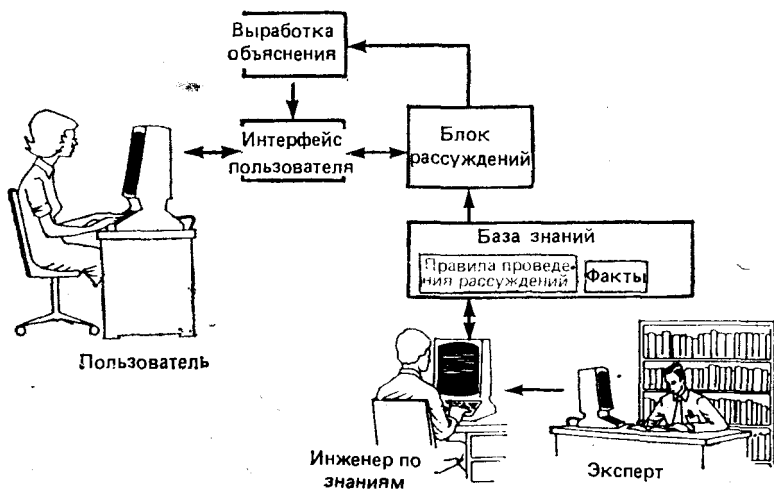


Рис. 10.23. Компоненты экспертной системы.

вать правила или фреймы для выработки заключения. Это обычно включает как схему получения логических выводов «от цели к фактам», так и прямое построение цепочки рассуждений «от фактов к цели». Алгоритмы, используемые при проведении таких рассуждений, очень часто представляют собой процедуры сравнения, в которых сравниваются истинность одного утверждения по отношению к истинности другого выражения. Примером служит алгоритм Рети (Rete), используемый в одной из широко известных экспертных систем (OPS-based, близкой к упоминавшейся выше системе R1). Этот алгоритм, работающий по схеме «от цели к фактам», сравнивает предпосылки в правилах продукции, обращенных к глобальной памяти.

Средства обоснования — это те средства, которые позволяют пользователю понять стратегию получения решений машиной, определяя, почему были получены конкретные выводы. Очень часто средства обоснования представляют модифицированный интерфейс естественного языка, позволяющий пользователю спросить машину:

«Как ты получила этот вывод?»

и получить ответ, показывающий эволюцию процесса рассуждений. Для системы, основанной на правилах, соответствующая реакция на поставленный вопрос могла бы выглядеть как список «запущенных в дело» правил в значительной мере подобно тому, как осуществляется трассировка в программах на Фортране. По мере того как размеры базы знаний увеличиваются и, следовательно, пропорционально растет ожидаемое число запусков правил, эта методика обоснования становится неудовлетворительной. В данном случае требуется такая система, которая могла бы подвести итоги процесса рассуждений, проводимого системой. Такое интерфейсное устройство дало бы особый выигрыш при разработке систем ПЕЯ, облегчая тем самым взаимодействие с экспертными системами, например, за счет использования архитектур рабочей области общего доступа.

Разработка «классических» экспертных систем обычно требует работы коллектива из трех человек: эксперта, инженера по знаниям и программиста символьных языков. Как было указано ранее, эксперт выдает знания и проводит экспертизу для системы, которая размещается в базе знаний системы. Инженер по знаниям получает знания от эксперта (обычно посредством интервью) и представляет их в виде модели. Эта процедура является сбором программируемых значений и она не зависит от любых знаний, получаемых системой либо от внешних датчиков, либо в процессе ее самообучения. Программист символьных языков получает информацию от инженера по знаниям и непосредственно вводит запрограммированные знания в систе-

му. Описанная технологическая схема требует больших затрат труда, и на разработку ранних вариантов экспертных систем уходили годы напряженной работы.

Такое разделение труда было существенно упрощено благодаря появлению средств разработки программных средств. И теперь инженер по знаниям и программист символьных языков очень часто совмещены в одном лице. Эти программные средства по существу представляют собой программы, содержащие все необходимое для получения экспертной системы. В такой инструментальной базе знаний снабжена всеми необходимыми элементами, но свободна от каких бы то ни было знаний. Предусмотрена такая возможность проведения логических выводов, позволяющая делать заключения сразу по мере заполнения базы знаний. Здесь уместна аналогия с электронной таблицей для микрокомпьютеров, т. е. описанные программные средства позволяют реализовать эквивалент пустой электронной таблицы. Программист имеет возможность вводить знания в базу знаний (подобно тому, как бухгалтер вводит цифры в электронную таблицу на дисплее), в результате чего он получает нестандартную, индивидуально удовлетворяющую его потребностям программу. Наконец, можно провести аналогию между совокупностью эвристических подходов для получения заключения и комбинацией ячеек электронной таблицы, используемой для организации новых входов. Подобно тому как электронные таблицы произвели революцию в использовании микрокомпьютеров, указанные средства уменьшили время, требуемое на разработку экспертных систем от нескольких лет до нескольких месяцев, в зависимости от уровня сложности задачи.

Основу экспертной системы составляет база знаний, поскольку, как было установлено, специализированные знания представляют очень важное дополнение к логике. Трудность процесса включения этих знаний в базу знаний является главным сдерживающим фактором даже в очень узких проблемных областях. Средства диалогового накопления знаний, такие как TEIRESIAS [19], доказали на практике свою полезность, позволяя эксперту выразить знания в формах, совместимых с базой знаний. Более сложные системы могут быть построены таким образом, что они сами будут способны выводить правила на основе представленных данных, как это было сделано в случае с MetaDendral [20]. Еще более успешно удастся создавать системы, которые могут обучиться управлять своими собственными стратегиями поиска, т. е. обучиться эвристическим методам, как это было в случае с EURISKO [21] и ACT [22]. Однако сами принципы обучения и адаптации до сих пор разработаны слабо, и в настоящее время большинство экспертных систем не являются обучающимися. Скорее они используют предварительно накопленные знания вместо обучения при ре-

шении задач на основе неудач и исправленных ошибок. Создание самообучающихся систем в том виде, как это описано в разд. 10.2 может значительно уменьшить время и усилия, затрачиваемые на разработку систем, основанных на знаниях.

Показателем совершенства интеллектуальных систем, и в особенности экспертных систем, является их способность приходить к заключению или решению несколькими цепочками рассуждений. Наличие данного свойства позволяет уменьшить пространство решений путем использования схем рассуждений либо отталкивающихся от цели и ориентированных на модель «от цели к фактам», либо соответствующих цепочке рассуждений «от фактов к цели». Это перспективное средство, хотя еще находится в начале своего развития, но уже сейчас позволяет анализировать задачу под разными «углами зрения». Следует заметить, что любой процесс, включающий одновременно несколько цепочек рассуждений, сильно зависит от наличия соответствующих структур параллельных вычислений. Это обусловлено ростом числа запросов к базе знаний, что в случае монопроцессорных систем приводит к непреодолимым трудностям. К сожалению, как будет показано в следующем разделе, наше понимание роли и методов реализации параллелизма в вычислениях является довольно ограниченным. Но потенциальная возможность использования параллельной обработки в символьных вычислениях является одним из факторов, делающих оптические вычисления особенно привлекательными.

Выяснив основные принципы работы экспертных систем, можно подробнее обсудить их возможности. Авторы [10] в своем руководстве по экспертным системам идентифицировали десять областей применения таких систем. Эти десять задач, показанные на рис. 10.24, предполагают, что экспертные системы пригодны для использования во всевозможных случаях принятия решений, от конструирования лазеров (в рамках заданных ограничений) до выделения источников поломок в оптических системах.

Рассмотрим подробнее п. 3 на рис. 10.24. Авторам кажется, что рассмотрение такого произвольно выбранного примера будет полезно для наглядного описания принципов работы экспертных систем, а также иллюстрирует идеи, обсуждаемые в предыдущих разделах.

В целях рассмотрения данного примера обратимся к системе, состоящей из лазера, узкополосного фильтра, нелинейного оптического устройства, фотодетектора и набора линз; схема этой системы показана на рис. 10.25.

В данной системе испускаемый лазерный луч поступает сначала в первую линзу, затем в фильтр, где световой пучок некоторым образом модифицируется. В данном примере будет предполагаться, что фильтр пропускает только свет, соответст-

вующий точной длине волны лазера, так что любые изменения в спектре выходного луча лазера приведут к уменьшению интенсивности света, проходящего через фильтр. Прошедший через фильтр свет затем пропускается через некоторое нелинейное оптическое устройство и выходной сигнал попадает на фотодетектор. Используя для описания данной системы последовательность фреймов на наиболее простом уровне, выделим в базе знаний элементы, показанные на рис. 10.26.

Путь, который проходит свет в системе, представлен словами От: и До:, указывающими на связи между различными фреймами в семантической сети. Изменения свойств света при его прохождении через систему представляются изменениями в конечном слове Выход:. Наконец признак Вход: накапливает знания о луче лазера, проходящем в системе. Предположим, что выходной сигнал детектора падает до нуля ($=$ Ток НУЛЬ) в языке Лисп) и перед пользователем возникает задача диагностики причин неисправности; экспертная система должна помочь ему в этом.

В приведенном примере система могла бы выполнять процедуры одним или двумя способами, работая либо по нисходя-

- (1) Интерпретация информации с датчиков;
- (2) Предсказание последствий в заданных ситуациях;
- (3) Диагностика сбоев;
- (4) Разработка объектов в рамках заданных ограничений;
- (5) Мероприятия по планированию;
- (6) Управление процессом наблюдения или сравнение результатов наблюдений с целью выявления уязвимых мест;
- (7) Поиск ошибок и выдача рекомендаций по преодолению сбоев;
- (8) Ремонт или выполнение плана рекомендованных операций;
- (9) Управление системами;
- (10) Команды, включающие диагностирование ошибок или средства регулирования действий обучаемого;

Рис. 10.24. Применение экспертных систем.

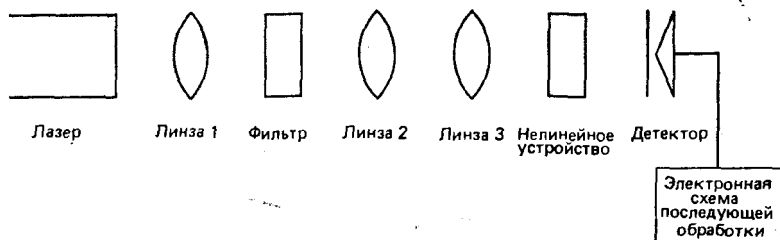


Рис. 10.25. Прототип оптической системы, использованной в приведенном примере экспертной системы.

шей схеме от «фотодетектора», либо по восходящей схеме от «лазера». В первом случае система могла бы предположить, что выходной сигнал отсутствует ввиду отсутствия входного сигнала; другими словами, отсутствует лазерный луч, достигающий детектора. Чтобы это было истиной, должно быть истинным одно из следующих утверждений: не исправен детектор, неисправно нелинейное устройство, поломка связана с какой-то из компонент, стоящей на пути светового пучка до уже упомянутых устройств.

Именно в этом месте в задачу может быть введена экспертиза, поскольку система может «знать», что:

Правило 1. Если выходной сигнал фотодетектора уменьшается до нуля, то фотодетектор не исправен.

Правило 2. Если в системе имеется нелинейное оптическое устройство, то его выходной сигнал чувствителен к правильности юстировки.

Применяя этот вид экспертизы к данной задаче, система может исключить фотодетектор и другие компоненты из числа причин возникновения задачи и выдвинуть гипотезу, что неисправно нелинейное устройство. Взаимодействуя с пользователем, система может принять решение об истинности начальной гипотезы. Это взаимодействие системы с пользователем может принять следующий вид:

Система: «Является ли входной сигнал Устройство равным Лазерное. излучение?».

Фрейм: Лазер	Фрейм: Линза. 1	Фрейм: Фильтр	Фрейм: Фильтр
От: Оптическая система	От: Лазер	От: Линза.1	От: Линза. 2
Вход: НУЛЬ	Вход: Лазерное излучение	Вход: Лазерное излучение	Вход: Лазерное излучение
Выход: Свет. 1	Выход: Свет.1	Выход: Свет.2	Выход: Свет. 2
До: Линза. 1	До: Фильтр	До: Линза.2	До: Линза.3

Фрейм: Линза.3	Фрейм: Нелинейное устройство	Фрейм: Фотодетектор
От: Линза.2	От: Линза.3	От: Нелинейное устройство
Вход: Лазерное излучение	Вход: Лазерное излучение	Вход: Лазерное излучение
Выход: Свет.2	Выход: Свет. 3	Выход: Текущая. запись
До: Нелинейное устройство	До: Фотодетектор	До: Оптическая Система

Рис. 10.26. Представление в виде фреймов оптической системы, приведенной на рис. 10.25,

Пользователь: «Да».

Располагая этой дополнительной информацией, система сделает заключение, что фактически причина поломки связана с устройством.

Для другой парадигмы система может предположить, что не исправен лазер и описать неисправность следующим образом. Если отсутствует Лазерное. излучение от Лазер, тогда отсутствует Лазерное. излучение на Линза. 1.

Линза. 1 : Вход = нуль;

Если Вход = Нуль, тогда Выход = Нуль;

Если отсутствует Лазерное. излучение от Линза. 1, тогда Лазерное. излучение не попадает в Фильтр;

и т. д. ...

Наконец, завершая:

Если отсутствует Лазерное. излучение от Лазер, тогда на Фотодетектор не попадает Лазерное. излучение;

Если Вход Фотодетектор = Нуль, тогда Выход Фотодетектор = Нуль.

Выработав гипотезу, что лазер имеет дефект, система могла бы ее проверить, спрашивая пользователя:

Система: Является ли выходной сигнал от Лазер к настоящему моменту времени равным Лазерное. излучение? (Соответствует ли Лазер: Выход = Лазерное. излучение?)

Если ответ положителен, то дефектными являются какие-то другие компоненты, стоящие на пути луча. Тогда система может предположить, что неисправен другой компонент, и продолжить процесс диагностики.

Приведенный выше пример является типичным для работы экспертной системы. Система приходит к заключению, основанному на истинности ряда элементов базы знаний, в любой конкретный момент времени. Гипотезы вырабатываются системой, а их справедливость проверяется через дополнительное взаимодействие с пользователем, поставляющим необходимую информацию о состоянии системы. Таким образом, экспертные системы могут функционировать как средство диагностики для различных пользователей.

Данный пример позволяет пользователю оценить трудности, связанные с включением знаний и экспертизы в систему ИИ. Приведенный случай является еще одной иллюстрацией, указывающей на необходимость компромисса между универсальностью системы и рабочими характеристиками. Одним из ограничений экспертных систем является узость области экспертизы, включенной в систему. Каждая экспертная система предназначена для определенных задач, и поэтому методики, разработанные для ее реализации, не могут быть использованы для других экспертных систем. В то же время увеличение размеров базы

знаний, эквивалентное расширению области экспертизы, приводит к комбинаторному «взрыву» числа возможных логических заключений и состояний машины. Имеются также и ограничения, связанные с аппаратным обеспечением, например размером рабочей памяти компьютера, куда могли бы накапливаться знания в произвольный момент времени.

Читатель может задать вопрос, почему увеличение размера рабочей памяти, приводящее таким образом в любой момент к возможности расширения активной базы знаний, хотя бы частично не облегчит проблему узости области экспертизы. Это действительно происходит, но при этом все же приходится столкнуться с трудностью организации и управления знаниями, и после всего этого возникает необходимость последовательно вводить знания по каналу в процессор. Это усложняется трудностями представления определенных типов знаний, таких как данные, зависящие от времени, данные, обладающие неопределенностью, зависящей от времени, а также знания о процессах и причинных связях. Типы представления, обсуждавшиеся в разд. 10.2.2, не адекватны для многих видов задач, поскольку они неспособны к соответствующему накоплению временных изменений или статических неопределенностей, связанных с этим знанием. Возвращаясь к нашему примеру поиска неисправности в оптической схеме, определим, что использование слов «обычно» и «типично» в правилах подразумевает неопределенность знания:

Правило 1: Если выходной сигнал детектора спадает до нуля, то обычно детектор неисправен.

Правило 2: Если в схеме имеется нелинейное оптическое устройство, то выходные сигналы устройства чувствительны к юстировке (установке оптических осей).

В настоящее время возможности представления этих типов знаний весьма ограничены.

В качестве примера рассмотрим проблемы, связанные с представлением объектов в пространстве, а также пространственными и временными соотношениями между ними. Эти проблемы широко распространены в области исследований технического зрения и они также возникают в экспертных системах в случае представления соотношений. Например, компьютеру трудно «понять», что световой пучок, падающий на линзу, как и свет, появляющийся с другой стороны, реально являются частями одного и того же пучка. Некоторые типы знаний, например зрительные образы, лучше хранить в несловесной форме. Дело в том, что их кодирование может потребовать очень большого числа компьютерных кодов. Однако в настоящее время не существует способа хранения графических знаний, позволяющего проводить модификацию, поиск и получение логических выводов.

Отметим также, что если бы можно было бы расширить область экспертизы, то все равно в обработке знаний существует ряд «узких мест», которые ограничивают производительность экспертных систем величинами, составляющими от 10 до 1000 логических заключений в секунду. Для сравнения заметим, что это приблизительно соответствует быстродействию от 1 до 100 Мегафлоп (операций с плавающей запятой) в цифровых вычислениях. Данный предел не является следствием только проблем аппаратного обеспечения; многое зависит от того, как программные средства организуют процессы поиска знаний и обоснования. Еще один аспект исследований экспертных систем заключается в нахождении эффективных способов разбиения на модули тех процедур, назначение которых накапливать и обрабатывать знания. В настоящее время фактически не существует способов разделения или разбиения на модули различных компонентов экспертных систем — средства обоснования и интерфейсы пользователя совместно используют одну и ту же память с базой знаний и механизмом вывода. Организуя, отслеживая и направляя все эти знания в монопроцессор, тем самым сильно ограничивают скорость, с которой могут выполняться символьные вычисления. Это является типичным примером ограниченности архитектуры фон Неймана, который будет обсуждаться в следующем разделе.

Все эти трудности в настоящее время ограничивают размер базы знаний и отсюда общую надежность системы. Вариант схемы, обладающей требуемыми характеристиками, показан на рис. 10.27, где компоненты экспертной системы разбиты на модули, но все еще взаимодействуют с системой на нескольких уровнях. Каждая компонента имеет возможность работать с отдельным процессором, но при этом в каждом параллельном процессе обработки осуществляется связь с другими обрабатываемыми элементами (посредством посылаемых сообщений). Такая структура системы могла бы обеспечить создание экспертных систем, занимающих большой объем памяти и имеющих более надежные средства логического вывода и более широкий диапазон областей применения.

Имея общее представление об основных функциональных возможностях символьных вычислений, можно увидеть, что все они направлены на то, чтобы улучшить понимание машинной области знаний, к которой относятся запросы пользователя. Это требует увеличения потока запросов к базе знаний системы. Кроме того, в данном случае важное значение имеют эвристический поиск и различные процессы проведения рассуждений. Однако все это в настоящее время представляет собой значительную трудность с точки зрения реализации.

Другое обстоятельство, заслуживающее внимания в связи с обсуждением указанного комплекса проблем, состоит в том,

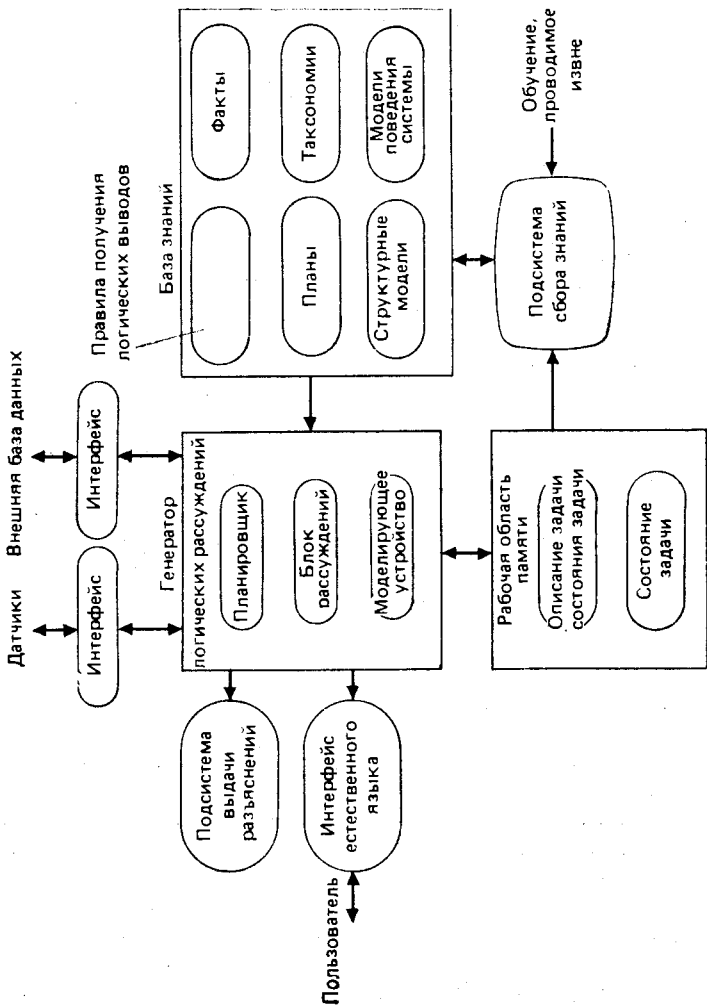


Рис. 10.27. Модульная параллельная экспертная система.

что виды обработки, используемые в системах ИИ, в большинстве случаев тесно привязаны к проблемным областям. Во многих случаях и команды, и методики использования памяти, и оценки, и метрики для сравнения, и процессы поиска тесно переплетаются и взаимосвязаны в последовательности процедур управления интеллектуальными системами. Следовательно, единственная возможность указать общие категории — это выявить степень пересечения методик обработки высокого уровня. Подобное можно видеть в сфере разработки инструментальных средств экспертных систем, где каждое программное средство обладает собственной формой представления данных (правила продукции, семантические сети, фреймы, сценарии), свои собственные парадигмы рассуждений (схемы поиска «от цели к фактам» и от «фактов к цели» и др.) и как результат имеет собственную структуру процедур сравнения и оценки. Хотя все эти программные средства в процессе работы могут взаимодействовать, но на практике различия между ними весьма велики.

Предметом рассмотрения в данном разделе являются проблемы, связанные с представлением знаний в системах ИИ. Любая работоспособная и в достаточной степени универсальная система требует использования целого ряда схем представления знаний. Проиллюстрировать данное утверждение можно на примере задач, использующих одновременно декларативные и процедурные знания. Таким образом, по мере перехода к решению более сложных и более общих проблем, системы будут работать со все большими по размеру базами знаний и увеличивающимся числом типов представления данных. Проблема будет заключаться в организации и управлении этими знаниями таким образом, чтобы это не приводило к возникновению дополнительных трудностей и уменьшению скорости обработки.

Согласно полученным оценкам, для того чтобы преодолеть существующие проблемы в системах ИИ, скорость вычислений в символьных компьютерах должна быть увеличена на несколько порядков по сравнению с уже достигнутой [23]. В то время как для решения этой задачи изучаются системы с высокой степенью параллелизма, ограниченность нашего современного понимания этой проблемы убеждает в том, что параллелизм сам по себе, в одиночку, не может обеспечить необходимого увеличения быстродействия. Использование альтернативных систем, таких как оптические вычислительные системы, может оказаться весьма перспективным, если соответствующие схемы кодирования данных и операций удастся реализовать оптическими методами.

В предыдущих разделах был выявлен основной круг задач символьных вычислений; особое внимание при этом уделялось потенциальной возможности применения оптических методик. Установлено, что наилучшим образом состыковать оптическую

обработку и вычислительные методы удается в задачах технического зрения и распознавания речи. Несмотря на то что в последнее время много говорят о возможности применения оптики для решения задач ИИ, остается целый ряд нерешенных проблем. К их числу относятся совместимое с оптикой представление данных, параллелизм (в целом в вычислениях и оптике), организация взаимодействия с памятью, оптимальная в конкретном случае конструкция оптоэлектронных интерфейсов, реализация оптимальных архитектур и технологии изготовления компонентов. Эти вопросы более подробно рассматриваются в следующем разделе, где описываются потенциально возможные архитектуры символьных вычислений.

10.4. Архитектуры символьных вычислений

10.4.1. Введение в архитектуры символьных вычислений

Преыдушие разделы подготовили необходимый фундамент для обсуждения архитектур символьных компьютеров. Обсуждавшиеся выше принципиальные характеристики символьных вычислений помогли осознать всю важность связей между элементами данных: связей между объектами и признаками в языке Лисп, связей между узлами в семантической сети, связей внутри и между фреймами и т. д. Это толкнуло специалистов в области компьютерной техники к исследованиям возможностей улучшения характеристик за счет использования таких архитектур компьютеров, для которых связь между узлами процессора отражала бы основные соотношения символьных вычислений. Такая идея вызвана частично опытом, накопленным при попытках увеличить производительность цифровых компьютеров путем приведения в соответствие их архитектуры со структурами алгоритмов и наоборот. Конечно, при этом следует помнить о важности поддержания гибкой архитектуры, которая может адаптироваться к изменяющимся связям между объектам; в противном случае в результате будут получены машины весьма специализированного назначения с ограниченными возможностями.

Сходство этих архитектур с нейронными системами создает кредит доверия к их значимости для символьной обработки. Мозг с его сравнительно медленно действующими компонентами (времена срабатывания нейронов составляют порядка миллисекунд) способен обрабатывать символьную информацию со скоростью по величине на несколько порядков бóльшей, чем для традиционных (фон-неймановских) архитектур компьютеров. Два различия между электронными и биологическими системами заключаются в способности организовывать соединения между компонентами и возможности произвольно комбини-

ровать операции с процессорами и памятью. Нейроны в мозге могут иметь более 10 000 синапсов (биологических соединений), в то время как их электронные аналоги — вентили в типичных случаях имеют лишь несколько соединений с другими вентилями. В области распределения ресурсов памяти фон-неймановские архитектуры характеризуются разделением между функциями процессора и памяти; пересылка информации между ними часто приводит к уменьшению быстродействия.

Авторы ни в коем случае не предполагают заниматься критикой фон-неймановских архитектур. В конце концов, в то время когда фон-неймановская структура процессор/память была предложена, она базировалась на ограничениях, накладываемых технологией того периода времени (конец 1940-х гг.). Кроме того, такие архитектуры доказали значительное превосходство над человеческим мозгом при выполнении цифровых операций. Хотя обсуждение оптимального соотношения между использованием символических и цифровых систем выходит за рамки данной главы, когда-нибудь в будущем системы смогут объединить в себе и символичные, и традиционные цифровые процессоры.

Компьютеры с высокой степенью организации межэлементных соединений и распределенной памятью были названы параллельными процессорами вследствие их способности выполнять совпадающие по времени операции. Поэтому, перед тем как перейти к обсуждению возможностей оптической параллельной обработки, авторы хотели бы познакомить читателя с основами и терминологией параллельной обработки, а также обсудить общие категории параллельной архитектуры.

10.4.2. Архитектуры параллельной обработки

Существуют многочисленные категории таксономии для классификации параллельных архитектур, но в данном обсуждении будут затронуты только те из них, которые, как представляется, будут полезны в контексте этой главы. Для более полного рассмотрения следует обратиться к монографии [24]. Прежде всего мы будем рассматривать вариант с временной параллельностью выполнения операций в отличие от пространственной параллельности. Первый из указанных вариантов наиболее часто принимает форму конвейерной обработки, которая представляет собой такое последовательное выполнение команд или операций, при котором начальные фазы последующих команд или операций вступают в действие до того, как завершены последние фазы предыдущих команд или операций. Системы с пространственной параллельностью сконструированы так, чтобы выполнять одновременно несколько частей задачи. Они состоят из двух или более элементов обработки и наиболее ча-

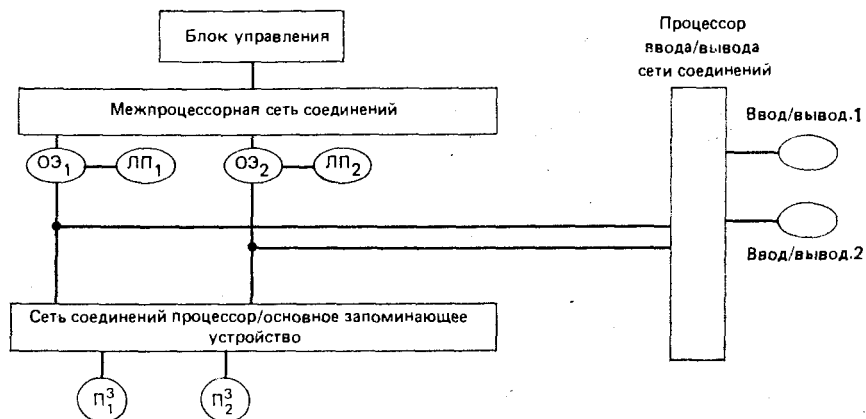
сто имеют приблизительно равные возможности, такие что каждый элемент содержит по меньшей мере арифметическое логическое устройство и набор регистров. Хотя взаимное расположение и способ соединения процессоров могут изменяться, схема на рис. 10.28 изображает общую концепцию их построения. Следует заметить, что сети межэлементных соединений, которые желательно делать программируемыми, играют главную роль в реализации параллельной обработки.

Классификация, представленная в [25] и изображенная на рис. 10.29, предлагает интересный вариант разбиения параллельных систем на категории исходя из числа процессоров и относительной степени сложности отдельного процессора. Традиционные монопроцессорные архитектуры изображены на рис. 10.29. По мере перехода к большему числу процессоров обычно происходит снижение степени сложности каждого отдельно взятого процессора, при этом следует стремиться достичь оптимального соотношения между ценой всей системы в целом и нарастанием ее сложности. Матрицы микрокомпьютеров представляют собой набор компьютеров, которые посылают сообщения друг другу посредством коммуникационной сети, как это изображено на рис. 10.30. Такие системы обычно характеризуются слабой связью элементов в противовес системам с сильной связью, т. е. отдельные компьютеры не делят между собой основную память и устройства ввода-вывода, хотя один компьютер всегда может привлечь ресурсы другого с помощью коммуникационной сети. Применение таких систем в символьных вычислениях, вероятно, будет происходить при решении задач, которые влекут за собой использование более чем одной базы знаний. Каждый процессор может работать с заданной частью задачи таким образом, чтобы свести к минимуму внутрипроцессорные связи.

Следующая категория, а именно вычислительные матрицы, являются системами, в которых обрабатывающие элементы выполняют операции со степенью сложности, сравнимой со сложностью операций умножения и сложения. Систематические матрицы, для организации которых процессоры соединяются в упорядоченные структуры, которые в свою очередь соответствуют потокам данных в вычислениях, охватывают большинство архитектур указанной категории. Однако в дальнейшем, вероятно, возникнут вычислительные матрицы более общего назначения; это произойдет по мере того, как сети межэлементных соединений станут более гибкими. Вопрос о систематических матрицах будет повторно рассмотрен при обсуждении гибридных оптоэлектронных систем (разд. 10.4.5).

Самые современные типы параллельных машин, оснащенные памятью, усиленной логикой и системами искусственных нейронных сетей, можно рассматривать как устройства «интел-

лектуальной памяти». Такие системы памяти могут значительно облегчить решение проблемы, которая возникает в фон-неймановской архитектуре из-за необходимости постоянно пересылать информацию в память и из нее. Они могли бы использоваться в качестве памяти для главного компьютера, т. е. пе-



ОЗ — обрабатывающие элементы

ЛП — локальное устройство памяти

П — модуль основного запоминающего устройства

Ввод/вывод — устройства ввода/вывода

Рис. 10.28. Блочная диаграмма универсального параллельного компьютера.

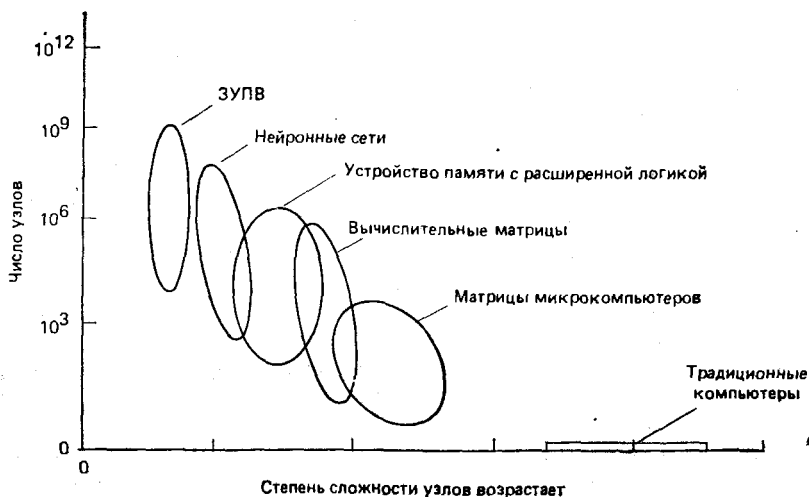
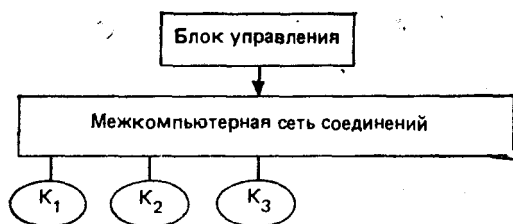


Рис. 10.29. Классификация параллельных процессоров в зависимости от сложности узлов.



К-компьютер (оснащенный устройством памяти и способный подсоединяться к устройствам ввода/вывода)

Рис. 10.30. Матрица микрокомпьютеров.

реводить часть процедур обработки в память во избежание снижения скорости работы, что связано с перемещением данных между изолированной памятью и блоками процессора. Каждый элемент (или узел) системы памяти, усиленной логикой, имеет объем несколько тысяч бит, набор регистров и некоторое количество связанной с ними логической информации, пригодной для работы с содержащейся на хранении информацией и способной указывать связи с другими элементами. Для случая искусственных нейронных систем уровень сложности узлов начинает приближаться к сложности переключающих элементов. Такие системы называют параллельными процессорами с разбиением на мелкие структурные элементы вследствие сравнительно низкой степени сложности отдельных элементов обработки. Такие элементы тесно взаимодействуют между собой и должны состоять по меньшей мере из нескольких тысяч элементов, чтобы достичь требуемой на практике производительности. Фактически на кульманах у конструкторов сейчас находятся варианты архитектур с уровнем разбиения, при котором число элементов достигает 1 миллиона.

Последний элемент, изображенный на рис. 10.29, или запоминающее устройство с произвольной выборкой (ЗУПВ) представлено для сравнения с системами с разбиением на мелкие структурные элементы. Вследствие отсутствия возможности соединений ЗУПВ не могут работать как мультипроцессоры.

Архитектуры с разбиением на мелкие структурные элементы и тесной связью между элементами представляют наибольший интерес для символьных вычислений вследствие того, что основную роль здесь играет их способность к организации межэлементных соединений, а не производительность. Например, каждый узел семантической сети может быть поставлен в соответствие отдельному узлу такой архитектуры и вся вычислительная мощность может быть направлена на установление и идентификацию типов связей. Некоторые из упомянутых выше машин, имеющие миллион процессорных элементов, под-

падают под категорию систем с памятью, усиленной логикой, и рассматриваются как основа для создания семантических сетей, состоящих из нескольких сотен тысяч связей; при этом возможности языка Лисп реализуются с помощью нескольких сотен тысяч соединяющих ячеек.

Иная популярная классификация параллельных систем связана с понятием организации архитектур с одним потоком команд и многими потоками данных (ОКМД); с многими потоками команд и многими потоками данных (МКМД). Имеются также две другие категории, а именно ОКОД и МКОД, определения которых для читателя должны быть очевидны. ОКОД представляет большинство существующих стандартных архитектур, а МКОД представляет концепцию, которой обычно уделялось очень мало внимания из-за ее сомнительной практической пригодности; следовательно, только архитектуры ОКМД и МКМД упоминались в контексте параллельной обработки. МКМД представляет полный параллелизм и как следствие является наиболее сложной из всех четырех категорий. Последняя схема требует индивидуального управления блоками каждого из процессоров и способности эффективно идентифицировать и распределять среди процессоров подмножество задач, находящихся на стадии обработки. Сети межэлементных соединений, допускающие взаимодействие между блоками, выделяют МКМД-системы среди систем, в которых задача может быть разделена на операции, выполняемые на нескольких ОКОД-машинах. Исследования МКМД-архитектур в большинстве случаев было ограничено системами со слабой связью, в основном вследствие ограниченности знаний о самих принципах параллельной обработки.

В ближайшем будущем наиболее значительный выигрыш в понимании параллельных методов, вероятно, будет получен от реализации ОКМД-архитектур, и, следовательно, именно это направление в настоящее время определяет наибольшую активность исследований в области параллельных архитектур. ОКМД означает, что не обязательно все процессоры выполняют один и тот же набор команд, но только то, что их характеризует одна и та же последовательность команд, и каждый процессор может быть включен в операции или выключен из них с помощью управляющего блока. Для символьных операций, которые наилучшим образом подходят для использования в больших машинах с разбиением на мелкие структурные элементы, применение архитектуры ОКМД может оказаться более оправданным, чем МКМД, из-за больших непроизводительных издержек, возникающих вследствие либо накопления целых множеств команд в каждом из процессоров, либо вследствие прохождения стандартных команд через межпроцессорную сеть соединений.

Вообще говоря, не существенно, к каким категориям отно-

сятся заданные архитектуры, поскольку все они сталкиваются с одной «захлестывающей» их проблемой — необходимостью создания сети соединений, которая могла бы эффективно выполнять перемещения сообщений в системе. Ограниченные возможности перемещения сообщений по сети соединений являются главным фактором, определяющим характеристики всей системы. Рис. 10.28 определяет три основные функциональные области, в которых важное значение имеют сети соединений: процессор — память, процессор — процессор и процессор — устройство ввода/вывода. И хотя рабочие характеристики в разных случаях существенно изменяются в зависимости от ограничений, накладываемых самими элементами, между которыми организуются соединения, существующие архитектуры применимы ко всем областям, а их выбор в прошлом происходил скорее на основании технических возможностей и приемлемой стоимости, чем функциональных различий. Последующее обсуждение сетей соединений соответственно уже не будет привязано к конкретной функциональной области.

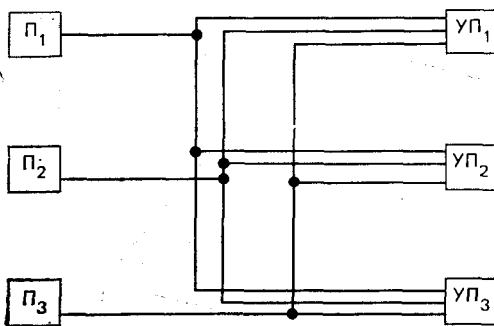
В архитектуре межэлементных соединений прежде всего стремятся продвинуться по пути реализации программируемых сетей. Это не является удивительным, поскольку производительность символьной архитектуры, особенно с мелкой структурой разбиения, определяется межэлементными соединениями. Именно увеличивая гибкость сетей и делая их программируемыми, разработчики проходят длинный путь в направлении увеличения общей вычислительной мощности. Более того, конструкторы систем обычно предпочитают несколько уменьшить быстродействие в обмен на устойчивость работы системы, заменяя жесткие межэлементные соединения на переключаемые, которые позволяют приспособить конфигурацию сети к имеющимся данным. Для символьных процессоров адаптивность имеет даже большее значение, чем для цифровых вследствие того, что топология структур данных обычно является неупорядоченной; следовательно, топология оптимальной сети не может быть определена заранее, до разработки схемы. Следует упомянуть, что программируемость соединений также важна с позиций устойчивости системы к неполадкам, поскольку при прохождении данных она позволяет перестраивать системы для обхода неисправных процессоров.

Сети соединений можно разделить на три общих класса: мультиплексные шины, многопортовые (порт — многоразрядный вход) компоненты и переключающие сети. Шина является простейшим, и, следовательно, наиболее популярным методом соединения, поскольку использует наименьшее число переключающих элементов. Однако конфликтные ситуации для этих переключателей в случае использования большого числа пользователей процессоров ограничивают степень их полезности

практически до уровня матриц мультимикропроцессоров со слабой связью, для которых конфликтные ситуации с распределением ресурсов шины не столь велики, как для систем с сильной связью.

Проблема конфликтных ситуаций может быть уменьшена путем создания компонент (например, запоминающих устройств) с более чем одним портом, а также за счет увеличения числа шин в системе, как это показано на рис. 10.31. Все это, естественно, увеличивает степень сложности компонентов.

В третьем из классов соединений используются сети переключателей, которые устанавливают каналы связи в системе. Наиболее общая сеть с обобщенным перекрестным переключателем способна устанавливать независимые связи между всеми компонентами, объединенными в сети, т. е. не существует разделения переключающих элементов между каналами и, следовательно, не возникает конфликтных ситуаций по поводу распределения ресурсов. Хотя такая сеть представляет собой наивысшее достижение с точки зрения мощности соединений, реализации на этой основе больших систем препятствует то, что цены на них оказываются слишком высокими, как и затраты на потребляемые мощности и защиту от перекрестных наводок. Универсальный перекрестный переключатель размерности $k \times m$ имеет k входных портов, предназначенных для соединения с k сигналами или узлами передачи сообщений, и m выходных портов для соединения с m принимающими узлами. Если бы каждый из узлов системы имел один передающий порт и один принимающий порт (или один двунаправленный порт), тогда $n \times n$ перекрестных линий позволили бы создать все возможные соединения между узлами и избавили бы от любых конфликтов по поводу ресурсов сети.



П — процессор

УП — устройство памяти

Рис. 10.31. Усиление возможностей системы межэлементных соединений посредством многопортовых устройств памяти.

Для простоты на рис. 10.32 изображены только перекрестные соединения размерности 3×3 . Следует заметить, что для приборной реализации перекрестных соединений размером $n \times n$ число переключателей составит n^2 , а число двунаправленных линий связи — $2n^2$. Для больших значений n реализация такой системы с помощью технологии электронных интегральных схем становится грандиозной задачей; особенно сложной является разработка $2n^2$ двунаправленных линий связи с эквивалентными полосами частот и приемлемым уровнем перекрестных наводок. Кроме того, электронные компоненты имеют очень ограниченный коэффициент разветвления по выходу, например для коэффициента разветвления по выходу электронного вентиля предел составляет приблизительно 10. Решение этой проблемы электронными методами заключалось в возврате к многоступенчатым переключающим сетям. Один из многих возможных вариантов показан на рис. 10.33, а, где каждый переключающий элемент имеет коэффициент объединения по входу, равный 2, и коэффициент разветвления по выходу, также равный 2. Сеть соединений является трехступенчатой, все переключатели в вертикальном столбце должны функционировать как одна ступень сети. Эта сеть межэлементных соединений 8×8 составлена из двенадцати 2×2 перекрестных переключателей, назначение которых изображено на рис. 10.33, б. В данном случае необходимо всего 48 переключателей ($12 \times 12 \times 2$) вместо 64, требую-

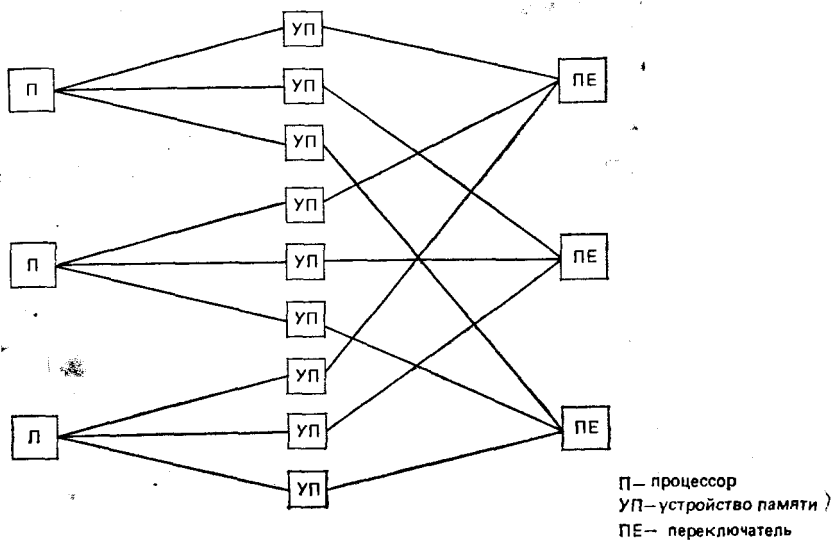


Рис. 10.32. Соединение процессор — память на основе перекрестного переключателя 3×3 .

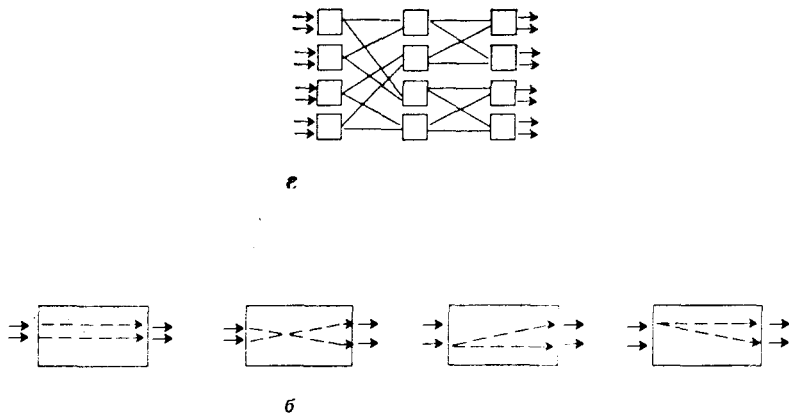


Рис. 10.33. Многоступенчатая сеть переключений:

а — топология основной сети, б — комбинация состояний переключателя.

щихся для перекрестной схемы 8×8 . Следует отметить, что, к сожалению, многоступенчатая конструкция оставляет возможность для появления конфликтных ситуаций. Для реализации соединений в перекрестных схемах малых размеров существуют многочисленные топологии (например, дерево, баньян, дельта, сетка и многие др.) и выбор обычно осуществляется между ценой и приемлемой частотой конфликтных ситуаций. Например, если степень конфликтных ситуаций определенной пятиступенчатой замкнутой сети является приемлемой, то схема соединений 1000×1000 потребует применения только 146 300 переключателей, а не 1 миллиона, который потребовался бы для перекрестного переключателя. Более подробное обсуждение топологии переключающих сетей представлено в книге [24].

Исходя из того что на практике применять универсальные перекрестные схемы нецелесообразно, разработчик сети должен решить, какая топология лучше удовлетворяет имеющейся структуре данных. Примером здесь может быть то, что основной топологией для соединений параллельных процессоров, предназначенных для обработки изображений низкого уровня, будет решетчатая структура, позволяющая осуществить соединения между ближайшими соседними элементами, так как подавляющее большинство операций низкого уровня использует изображения из смежных ячеек. Но по мере перехода к операциям более высокого уровня все более важными становятся глобальные связи между данными, требуя соответственно более глобальных топологий межэлементных соединений.

10.4.3. Оптическая реализация мультипроцессорных архитектур

Любой подход к рассмотрению оптических архитектур требует серьезных размышлений на тему, а что вообще можно реализовать на основе существующих технологий, и в частности на базе электронных схем со сверхвысокой степенью интеграции. Оптические вычисления не будут создавать серьезной конкуренции электронным вычислениям, если не дадут существенного (на несколько порядков) улучшения характеристик по ряду критических параметров. Следовательно, отправной точкой при рассмотрении применений оптики в символьных вычислениях должно стать выявление круга задач, решаемых с помощью электронных систем.

Вполне закономерным является тот факт, что относительная слабость и сила электронных и оптических методов прослеживаются тем или иным путем уже в самих основах физики междуэлектронных или междуфотонных взаимодействий. По этому поводу заметим, что между электронами существует сильное взаимодействие, в то время как между фотонами — слабое. Отсюда следует, что электроны хорошо подходят для использования в переключающих операциях, столь важных в вычислениях, а фотоны хороши для осуществления связи между переключателями, создавая связи, свободные от вредных эффектов взаимодействия, создающих перекрестные наводки и емкостную паразитную нагрузку. Однако соглашаться с такими утверждениями было бы преждевременно вследствие существования квантовых потерь, сопровождающих как преобразование электрона в фотон, так и фотона в электрон. Исследования и разработка направлены на поиск путей замены длинных внутрикомпьютерных соединений оптическими каналами связи, поскольку именно длинные соединения создают для электронных устройств жесткие проблемы с выделением мощности, быстродействием и занимаемым объемом для электронных устройств [26]. Это, естественно, слишком далеко от реализации оптикой всех ее возможных преимуществ в многопроцессорных архитектурах, пригодных для символьных вычислений.

Рассмотрим качественную схему соотношения объема электронных переключений и оптической связи, представленную на рис. 10.34 в виде квадрата, для которого стороны представляют континуум состояний между экстремальными состояниями, изображенными по углам. Верхний левый угол представляет чисто электронные системы, в то время как нижний правый угол — чисто оптические системы. Так как перемещение в направлении левого нижнего угла выходит за рамки рассматриваемой проблемы, внимание будет уделено состояниям, расположенным вверху и справа. При рассмотрении вычислительных систем, для которых переключение представляет более важную функ-

цию, компромисс между оптикой и электроникой, как видно, соответствует области схемы, расположенной где-то ближе к верхнему краю схемы, т. е. соответствует чисто электронным переключателям с некоторыми оптическими связями. Однако символьные методы обработки уделяют большое внимание организации связей, как уже было указано ранее в данной главе. Перенос акцента с переключений в архитектуре с разбиением на мелкие структурные элементы на организацию связей системы с сильной связью приводит к рассмотрению архитектур, для которых связь осуществляется оптически, и только некоторые из переключений выполняются электронными методами. Эта та категория гибридных электронно-оптических архитектур, которая, как уверены авторы, будет оказывать значительное воздействие на символьные вычисления.

Ближе к верхнему краю изображенного на рис. 10.34 континуума состояний располагались бы те архитектуры, в которых применяются оптические переключения при реконфигурации соединений, но при этом для логических операций используются электронные переключения. Как упомянуто выше, оптика докажет особую значимость при реализации длинных межэлементных соединений вследствие существования ряда присущих длинным электронным соединениям отрицательных свойств, касающихся потребления мощности, быстродействия и занимаемых объемов. Пример архитектуры с аналогичным соотношением оптических и электронных методов показан на рис. 10.35. Для простоты изображены только две из многих возможных

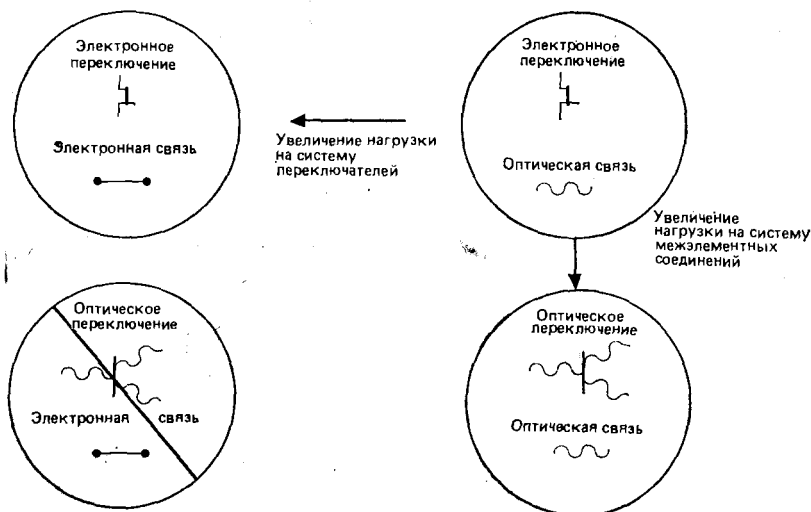


Рис. 10.34. Соотношение электронных и оптических вычислений.

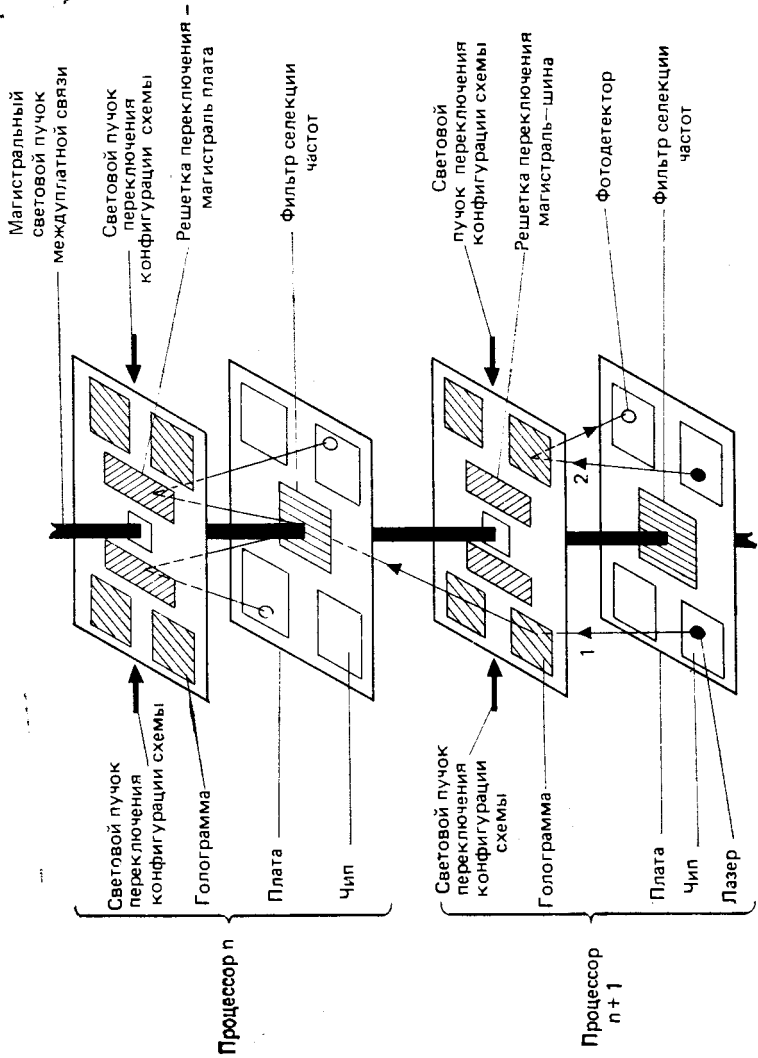


Рис. 10.35. Гибридная оптоэлектронная мультипроцессорная архитектура.

плат, и на каждой из плат показаны лишь четыре чипа. Если бы это был процессор с разбиением на мелкую структуру, то каждый чип сам по себе содержал бы много обрабатывающих элементов. Например, один такой электронный символьный компьютер, разрабатываемый в настоящее время и получивший название машины соединений [27], составлен из большой матрицы печатных плат, каждая из которых содержит 512 обрабатывающих элементов, равномерно распределенных по 32 чипам (т. е. на один чип приходится шестнадцать обрабатывающих элементов).

Каждая плата на рис. 10.35 содержит четыре оптоэлектронных чипа и один частотно-селективный фильтр (голограмму). Между каждыми из двух соседних плат расположена планарная матрица перестраиваемых дифракционных решеток, выполняющих большинство переключающих операций, необходимых в процессе соединения элементов. В этой конкретной архитектуре используется мультиплексирование со спектральным разделением каналов для оптических информационных потоков, направленных к соответствующим платам. Эту операцию осуществляет поток света, обозначенный цифрой 1. Голограмма, расположенная непосредственно над пропускающим свет чипом, направляет поток света в центр следующей платы, где тот накладывается на основной поток света, проходящий через все платы системы. По мере достижения требуемой платы частотно-селективный фильтр дифрагирует свет на голограмму связи магистраль — плата, направляющую поток света по его окончательному назначению.

Междуплатные и междучиповые соединения следует осуществлять с помощью плоской матрицы голограмм, расположенной над платой, что изображается пучком света, обозначенным цифрой 2. Логические процедуры управления большим числом мультиплексированных пучков в данном случае не обсуждаются, однако следует заметить, что оптические переключения наиболее вероятно будут осуществляться с помощью нелинейного смещения частот. Например, четырехволновое смещение частот может быть использовано для получения голограмм, которые могут быть быстро модифицированы, чтобы дать возможность перестроить межэлементные соединения. Переключающие пучки света, показанные на рис. 10.35, содержали бы информацию, нужную для управления голографическими решетками. Следует заметить, что некоторые из переключающих процедур в таких архитектурах могут выполняться с помощью мультиплексирования, а различные голограммы могут использоваться как пассивные решетки, выполняющие селекцию различных длин волн.

Реализация представленной выше гибкой схемы соединений, основанной на переключении световых пучков в свободном пространстве, позволит избежать одну из наиболее трудных проб-

лем для электронных символьных вычислений, заключающуюся в сложности реализации перестраиваемых соединений. В настоящее время возможности электроники определяются наличием проводов, используемых для всех соединений, что ограничивает как возможности изменения конфигурации системы, так и коэффициент разветвления по выходу. Это устанавливает жесткие ограничения на архитектуру переключающей сети, которая, кроме того, усложняется необходимостью обеспечения интенсивных переключений подобно тому, как это было обнаружено для задач ИИ. Спектр оптических сетей, обладающих намного большей гибкостью, здесь не будет рассматриваться; однако заинтересованному читателю следует обратиться к работе [29].

10.4.4. Чисто оптические архитектуры

По мере перемещения в правый нижний угол классификационной схемы на рис. 10.34 доля оптических элементов увеличивается до тех пор, пока не получается чисто оптическая архитектура. Пример оптического компьютера с разбиением на мелкие структурные элементы и сильной связью между элементами показан на рис. 10.36. Хотя никто еще не построил подобный компьютер, технически возможно создать систему, состоящую из 1 миллиона параллельных каналов. Это отнюдь не означает, что система включала бы конфигурацию обязательно из 1 миллиона узлов, так как такая конфигурация не подразумевает, что планарная матрица логических элементов, обозначенная как матрица вентиля, имела бы именно один логический элемент на канал. Вместо этого несколько логических элементов следовало бы соединить посредством среды межэлементных соединений, что позволило бы образовать элемент процессора. Например, квадратная матрица $n \times n$ логических элементов (вентилей) может содержать блок арифметической логики, несколько регистров и, возможно, несколько устройств кэш-памяти (быстродействующей буферной памяти большой емкости). Пример структуры указанного типа представлен на рис. 10.37, где для отдельных элементов двумерного ПМС были обозначены основные функции, присущие элементам вычислительной обработки. Принимая n равным 5 (25 логических элементов на процессор), в итоге получаем, что в машине должно быть 40 000 узлов, что составляет достаточно большую величину, чтобы такое устройство имело смысл использовать в качестве символьного оптического компьютера, реализующего символьные вычисления.

Ввод сигнала в оптический компьютер можно было бы осуществлять посредством матрицы лазерных диодов, адресуемых независимо друг от друга, или двумерных ПМС. Матрица диодов обладала бы намного большими скоростями модуля-

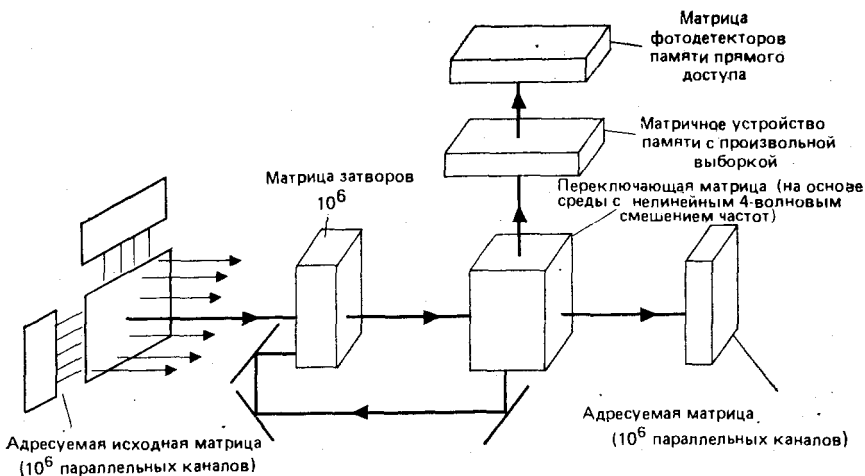


Рис. 10.36. Чисто оптическая мультипроцессорная архитектура.

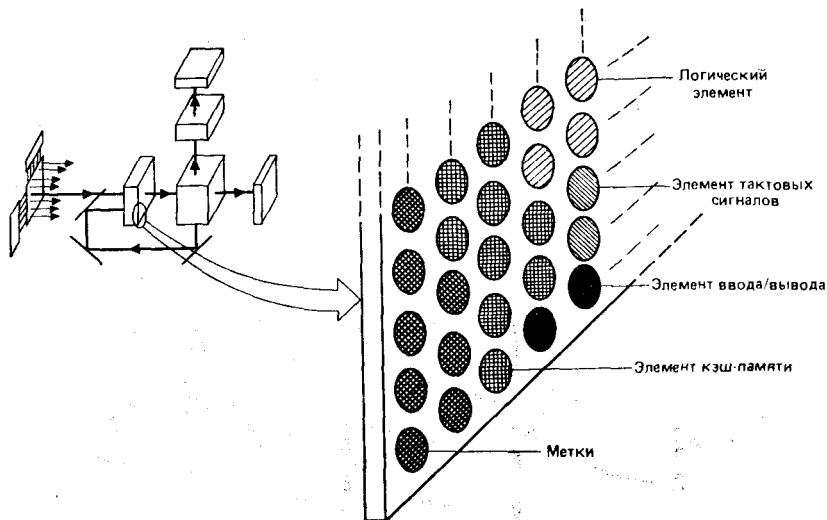


Рис. 10.37. Чисто оптический обрабатывающий элемент.

ции, чем ПМС, но требовала бы намного более сложной схемы, особенно если условия работы требуют однородности излучения по всей матрице. Если входной сигнал уже имеет вид двумерного изображения, которое могло бы выводиться из процессора технического зрения, то использование входного устройства не является обязательным, а целесообразность его использования зависит от степени совместимости двух процессоров.

Матрица логических элементов могла бы представлять собой двумерный ПМС, обладающий нелинейной характеристикой, или матрицу оптических бистабильных переключателей. Последнее устройство в конечном счете позволило бы получить намного большие скорости переключения, однако существующие виды оптических бистабильных переключателей потребовали бы слишком высоких уровней мощности.

В настоящее время разрабатываются более совершенные нелинейные материалы, позволяющие создать более пригодные на практике оптические бистабильные устройства. Элементы соединений, вероятно, будут использовать эффект смещения частот в нелинейных оптических средах и будут аналогичны по сути устройствам, описанным ранее для гибридных архитектур. Однако из-за намного большего числа каналов, с которыми необходимо работать, переключения должны выполняться многоступенчатым образом, когда ряд параллельных плоских голографических матриц, работающих в реальном времени, выполнял бы процедуры аналогично тому, как это показано на рис. 10.38. Заметим, что для простоты все три основные функции межэлементных соединений процессор — процессор, процессор — память, процессор — устройство ввода/вывода объединены в один блок, но они могли бы быть реализованы тремя независимыми устройствами.

С точки зрения технологии главную трудность представляло бы создание детектора. В большинстве случаев предпочтение было бы отдано устройству с 1 миллионом каналов, где каждый канал работает на частоте 1 МГц (проектируемое быстродей-

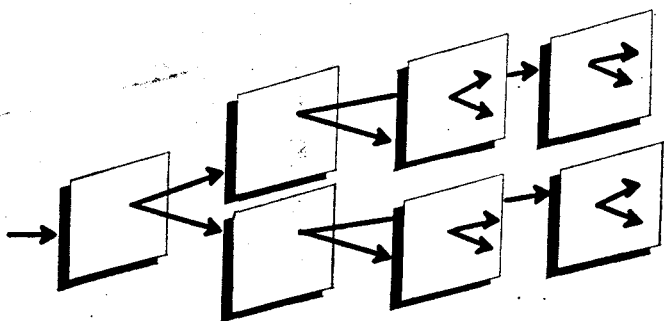


Рис. 10.38. Многоступенчатая оптическая сеть межэлементных соединений.

ствие для двумерных ПМС). Однако для наиболее практических конструкций процессоров потребности были бы намного ниже. Если бы проблемная область требовала, скажем, выполнения 100 итераций или более (например, семантическая сеть ведет поиск на глубину по меньшей мере 100 шагов), выходной сигнал требовалось бы выводить только один раз за каждые 100 мкс. Это снижает требования к пропускной способности детектора до 10^{10} бит в секунду (бит/с); последняя величина более согласовывается с ожидаемыми возможностями устройств на основе GaAs. В качестве примера можно было бы рассмотреть случай, где каждый процессор, как это было ранее, состоит из блока $n \times n$ каналов. Полагая, что $n=5$ и каждый процессор имеет именно один выходной канал, пропускная способность детектора составила бы 4×10^{10} бит/с. Определенная комбинация этих двух разработок предъявила бы такие требования к детектору, которые хорошо укладывались бы в реально выполнимые технические требования.

Последняя из основных компонент этой оптоэлектронной архитектуры — это устройство памяти. В электронных компьютерах одна из целей состоит в размещении большей части памяти вместе с логическими элементами. В области оптических компьютеров это не является ни необходимым, ни желательным, потому что задержки в передаче сообщений сильно снижены по сравнению с электронными устройствами. Таким образом, на рис. 10.36 основное устройство памяти изображается как единый блок, равно используемый всеми процессорами посредством двунаправленных каналов связи.

Другой важной областью применения оптики являются многопортовые устройства памяти. Фактически использование нескольких длин волн могло бы для любого заданного участка памяти обеспечить считывание при одновременном использовании большого числа каналов. Это позволило бы избежать необходимости использования сложных цепей выделения интересующей информации. Например, голографические решетки могли бы быть использованы для демультимплексирования большого числа наложенных друг на друга отраженных сигналов с различными длинами волн, отраженных от заданного пятна на оптическом диске; кроме того, могли бы быть использованы голографические элементы памяти, которые пространственно разделили бы различные считанные длины волн. Способ, которым это могло бы быть реализовано, показан схематически на рис. 10.39, где большое число пучков света могло бы одновременно использоваться для адресации оптического диска.

Другим привлекательным свойством применения нескольких длин волн в оптических вычислениях является то, что управление переключением осуществляет сам луч, несущий информацию, и не требуется предусматривать отдельный вход

для управляющего луча. В последнем случае это существенно увеличивало бы сложность операций управления компьютером. Напротив, использование целого ряда длин волн делает в большей мере параллельными операции, используемые для маршрутизации сообщения, в которых начальные биты в общем потоке битов сообщения содержат информацию об адресе, используемую каждым переключателем, с которым сталкивается сообщение по мере прохождения по сети.

Вычислительная мощность обработки с чисто оптической архитектурой могла бы быть увеличена посредством применения конвейерной обработки. Этого можно было бы достичь, копируя матрицу логических элементов подобно тому, как это показано на рис. 10.40. Конвейерная обработка была бы полез-

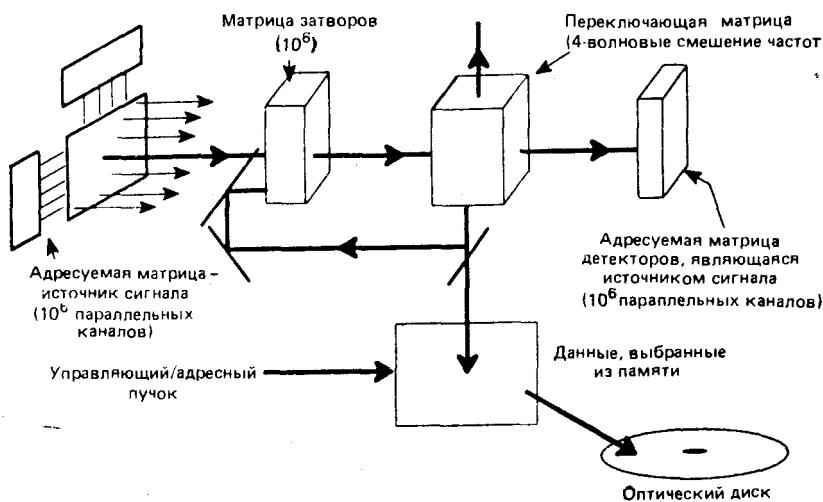


Рис. 10.39. Интерфейсное устройство на основе оптического диска для чисто оптической системы.

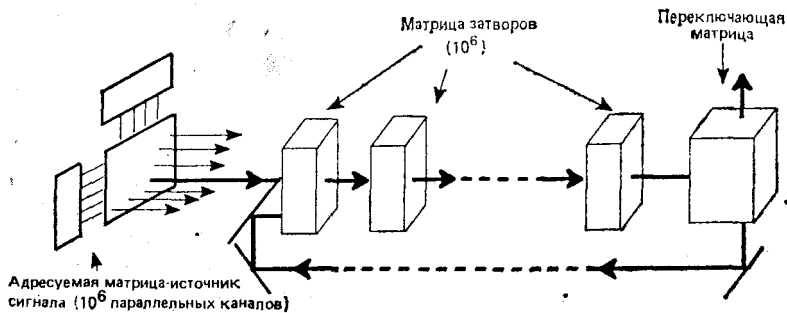


Рис. 10.40. Матрица оптических затворов для конвейерной схемы обработки,

на для многомерных задач, таких как обработка в системах технического зрения, связанных с обработкой зависящих от времени трехмерных изображений (например, каждая плоскость могла бы выполнять роль различной глубины изображения).

К настоящему моменту читателю следовало составить впечатление о типах архитектур, необходимых для символьных вычислений, а также о нескольких способах выполнения их по своей сути аналоговыми оптическими устройствами. Должно быть ясно, что даже эти «чисто оптические» структуры в некотором смысле являются гибридными оптоэлектронными архитектурами. В частности, электроника должна была бы использоваться для интерфейсных устройств связи с пользователем, для цифровых контроллеров и т. д., тогда как оптика должна была бы использоваться для проведения символьной обработки. В отношении схемы на рис. 10.34 можно было бы ожидать, что имеется спектр уровней, где могут возникать оптоэлектронные интерфейсы. Этот спектр возможностей архитектур простирается от чисто электронных до гибридных оптоэлектронных систем и приводит к рассмотрению других потенциальных возможностей использования оптики для задач символьных вычислений. Некоторые из этих возможностей рассматриваются в следующем разделе.

10.4.5. Гибридные оптоэлектронные системы

Имеется несколько уровней, на которых оптика может эффективно сочетаться с электроникой. На рис. 10.41 изображена иерархия функций в гибридной системе, простирающаяся от применения оптического процессора в чисто оптических системах, описанных в предыдущем разделе, до использования оптики только в периферийных устройствах, таких как оптические диски для хранения информации. Различия состоят в степени связи между электронным процессором и оптической системой и в объеме вычислений, выполняемых оптически. Далее предполагается, что электронная система является базовым процессором, а оптическая система соединяется с ним посредством одной из шин системы.

На нижнем уровне (рис. 10.41) оптическая система полностью заменила бы электронную систему на уровне процессора и, следовательно, выполняла бы большинство вычислений и обеспечила бы сильную связь с памятью электронной системы. Согласно терминологии разд. 10.4.2, такую гибридную схему можно было бы назвать системой с сильной связью. На следующем уровне имеем структуру, где оптическая система вычисляет некоторые из простейших операций (умножение, сложение, вычитание и т. д.), в то время как электроника выполняет

другие операции. Примером этого могло бы быть устройство оптического сопоставления с образцом, связанное с шиной данных Лисп-машины, где оптическая система выполняла бы все операции сравнения, оставляя электронике вычисления других простейших функций. Это система также является системой с сильной связью, поскольку процессоры распределяют реально имеющиеся ресурсы памяти, и по своей сути аналогична концепции оптического сопроцессора. Ширина полосы частот сети соединений должна быть очень высокой и весьма приблизительно равняться увеличенному в несколько раз объему выборки, деленному на время пересылки.

Ускорителями называются процессоры, которые резко увеличивают производительность системы конкретной функции,

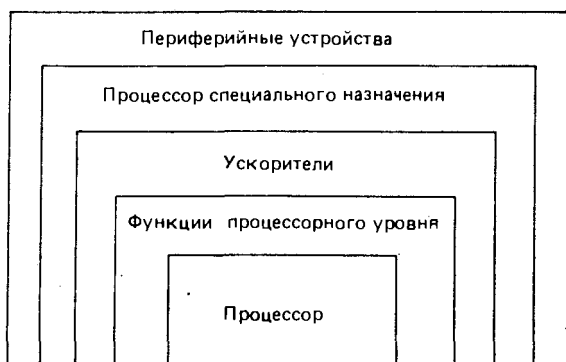


Рис. 10.41. Иерархия функций в гибридной системе.

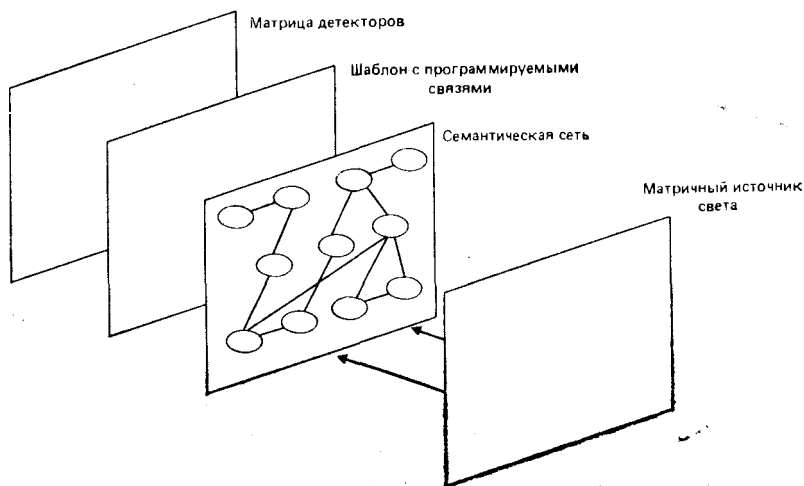


Рис. 10.42. Параллельная обработка в семантической сети.

такой как скалярное произведение, корреляционная функция или инициация правил. Как стало ясно на основании разд. 10.3, скалярное внутреннее произведение играет главную роль в таких процедурах обработки в системах ИИ, как сравнение признаков, сопоставление с образцом, корреляционная обработка на низком уровне, поиск в базе знаний и получение логических выводов на высоком уровне. Позднее в данном подразделе будет рассмотрено, как внутреннее произведение может применяться при обработке правил типа «если... то...».

Ускорители в свое время оказали заметное влияние на микропроцессорные цифровые вычисления, помогая свести к минимуму влияние явно вызывающих сомнения «узких мест». Удивительно, что они еще не получили широкого применения в вычислениях или в мультипроцессорных системах; оптика могла бы помочь ускорить этот процесс. Как показано на рис. 10.42, оптический компьютер мог бы быть соединен с шиной данных системы, но она в данном случае не распределяет ресурсы памяти в электронной системе. Это является примером гибридных систем с более слабой связью, где базовая машина и ускоритель расположены достаточно близко, но не обязательно в одной и той же стойке.

Процессоры специальных функций (ПСФ), как и предполагает их название, позволяют получить взамен универсальности улучшенные характеристики, максимально увеличивая производительность для конкретных функций. В типичных случаях они являются исключительно специализированными компьютерами с ограниченной способностью к программированию, ограниченной памятью и минимальными требованиями к интерфейсу. В качестве отдельных вычислительных блоков ПСФ соединяются с базовым компьютером посредством сети, оптического волокна или каких-либо других широкополосных сред. Их также называют вычислительными матрицами, и они успешно использовались для выделения признаков в системах технического зрения низкого уровня и речи, в качестве процессоров дисплея в компьютерной графике и как матричные процессоры в вычислениях быстрого преобразования Фурье. Наиболее популярные ПСФ являются систолическими матрицами, для которых имеется богатая экспериментальная база и также велико число возможных оптических вариантов реализации.

В оставшейся части раздела авторам хотелось бы остановиться на двух приведенных выше примерах, а именно использовании оптических вычислений для реализации ускорителя и процессора специального назначения. В случае ускорителя будет использоваться пример обработки внутреннего произведения для правил типа «если... то...». В качестве процессора специального назначения рассмотрим случай применения систолической матрицы для обработки семантической сети.

В символьных вычислениях центральное место занимает операция вычисления внутреннего произведения, эквивалентная умножению составляющих элементов на вектор (векторное умножение), на матрицу (умножение матрицы на матрицу) или на корреляционную функцию. В предыдущих разделах была установлена общность процедур вычисления внутреннего произведения для большого числа алгоритмов из области цифровых вычислений. В одном типичном представлении символьных вычислений отношения знаний выражаются в терминах логического сопоставления с образцом, процедура которого определяется поиском соглашения по предпосылке-условию (с левой стороны) соотношения «если $[A]$, тогда $[B]$ » (см. разд. 10.3.5). Здесь $[A]$ является подпространством N -мерного векторного пространства:

$$[A] = \{\text{данные/объекты, принадлежащие классу } A\},$$

в котором рассматриваются M векторов, где $M \leq N$

$$\{a_i(k) \mid a_i \in [A], \quad k = 1, \dots, M\}.$$

Принадлежность к этому подпространству может быть проверена с помощью операций над простыми функциями. Например, нулевой функционал подпространства единственным образом выражается через вектор a_i , являющийся ортогональным подпространству $a(k)$. Тогда внутреннее произведение

$$\langle a_i \mid a_j(k) \rangle = 0 \quad \text{для всех } a_i(k) \in [A].$$

Таким образом, вычисление или в данном случае запуск правила между элементами знания в этом представлении и некоторым соответствующим функционалом могут быть выражены с помощью операции вычисления внутреннего произведения. Более того, эта логическая структура могла бы быть использована либо в восходящей схеме проведения рассуждений, ориентированной на цель, так как в каждом случае предпосылки в правиле продукции «если... то...» должны быть удовлетворены в процессе рассуждений. Это также справедливо для представлений, основанных на фреймах, поскольку в этом случае каждый элемент знания представляет двумерную матрицу информации, обрабатываемую целиком, а $\{a_i\}$ может иметь вид матричных компонентов. Для каждого запуска правил операция сравнения может очень эффективно вычисляться в оптической системе, облегчая серьезные затруднения, возникающие в ряде электронных систем ИИ.

Внутреннее произведение является не единственной операцией, где оптика может иметь большое значение. Систематические матрицы, реализованные в нескольких вариантах оптических устройств, как было показано в [30—32], имели опреде-

ленные отображения на сети графа сигнал — поток. Это подразумевает, что задачи, рассмотренные с точки зрения теории графов или сразу основанные на моделях теории графов, могут иметь непосредственные отображения на хорошо известные топологии систолических матриц. В символных вычислениях анализ теории графов был применен для разработки соотношений между искомыми объектами и их свойствами; это исследование привело к разработке представлений о семантической сети.

Для наглядности семантическая сеть может рассматриваться как набор узлов, представляющих символы, соединенные связями, представляющими отношения. Наиболее фундаментальным соотношением между символами является связь «Является А»; другими типами соотношений могли бы быть «В Расположении», «Член-Множество», «Часть». Такие отношения определяют область и зависят от таксономии исследуемой задачи. В этом представлении основным вопросом, общим для систем символных вычислений было бы «Является ли элемент [A] элементом, [B]? Если предположить, что все связи между общим классом узлов [S] удастся представить связями «Является [A]», тогда этот вопрос сводится к задаче «Для [A], принадлежащего [S]»,

«Является ли [B] также принадлежащим [S]?» и

«Существует ли связь между [A] и [B]?»

В традиционной архитектуре этот второй вопрос включал бы обширный последовательный поиск по всей памяти путей, исходящих из [A]. Однако в систолической системе вывод включает только число шагов между [A] или [B], или до края матрицы (для [B], не являющегося элементом [S]); это происходит потому, что поиск происходил бы вдоль всех ветвей одновременно.

Чтобы найти соответствие поставленной задачи с архитектурой, являющейся потенциально систолической, можно изобразить каждый узел включенным в процессор или ячейку, а связи с соседними узлами описать с помощью уже существующих топологических связей в матрице. Здесь матрица представляет трехмерную конструкцию, где третьим измерением является время. Узел [A] может быть обозначен как элемент, представляющий интерес, а окончательный поиск в направлении [B] может выглядеть как поток битов между узлами, распространяющийся в каждый момент времени. Это показано схематически на рис. 10.42. В то время как это все еще требует некоторых тестов для проверки условия сравнения, набор задач семантической сети должен отображаться на систолические матричные системы.

В заключение заметим, что оптика предлагает несколько универсальных по своим возможностям архитектур, обсуждав-

шихся выше и обладающих большими перспективами для реализации символьных вычислений. Наиболее важными из возможностей являются реализация быстродействующих глобальных соединений, перестраиваемость соединений, создание элементов с высокими коэффициентами разветвления по выходу и многопортовых компонент для параллельной обработки.

Авторы хотели бы выразить благодарность тем из коллег, кто затратил свое личное время на совместную с ними работу с целью улучшения качества данной рукописи. Доброжелателей было так много, что поименное их перечисление выльется в очень длинный список. Однако некоторые из них заслуживают особой благодарности. В начале списка стоят д-ра Р. А. Атали, Т. А. Блэкстон, Р. А. Гизи, Л. Х. Рикер, чьи глубокие и своевременные советы были особенно ценными, и многие из их мыслей, объяснений и интерпретаций нашли отражение в рукописи. Далее идут: Дж. Л. Перри — за его помощь в вопросах понимания изображений и за согласие предоставить авторам некоторые материалы из его коллекции компьютерно-синтезированных изображений, д-р Дж. Б. Джилмер — за его идеи по обработке семантических сетей и использование систолических архитектур, Т. С. Старк — за его глоссарий и помощь в построении моста через «пропасть» в технологии между оптикой и символьными вычислениями и д-р С. Т. Батлер — за помощь авторам по ряду проблем, обеспечивающую им возможность выделить время для более тщательной проработки своих идей.

Глоссарий

Аналогические знания (или знания по аналогии): Знания, представляемые в виде аналогий. Примером этого являются образцы звуков, представляющие слова в системах обработки речи, или представления изображений двумерной матрицей чисел, соответствующих градациям шкалы полутонов.

Архитектура: Способ организации отдельных элементов в компьютере.

Атом: Символ (либо константа, либо переменная), используемый для обозначения объекта в программах на языке Лисп.

Атрибут: Часть описания объекта, содержащаяся во фрейме. Атрибуты обычно указывают такие характеристики, как цвет, размер и величина. То же самое, что слот.

База знаний: Хранилище знаний в компьютерных системах искусственного интеллекта.

Вероятность: Мера доверия пользователя к справедливости предположения, гипотезы или правила выполнения логических выводов.

Восходящая схема решения: Метод решения задач, осуществляемый из начального состояния до некоторого желаемого состояния. (См. также «Вывод „от фактов к цели”».)

Вывод «от фактов к цели»: Процедура решения в системах, основанных на знаниях. Процедура исходит из данных, задаваемых пользователем, и направлена на поиск решения, адекватно удовлетворяющего начальным данным. (См. также «Прямой логический вывод».)

Вывод «от цели к фактам»: Рекурсивная процедура решения задачи в системах, основанных на знаниях. Последовательность логических выводов, основывающихся на заранее поставленной цели и осуществляемых от состояния цели до одного из более

ранних состояний. Если получение последовательности выводов оказывается невозможным, то состояние цели изменяют. Если удастся получить последовательность выводов, то начальное состояние (или состояния) становится состоянием цели, и процедура начинается снова.

Декларативные знания: Знания, состоящие из фактов или утверждений.

Дерево решений: Представление возможных решений в задаче искусственного интеллекта в виде дерева с вершинами, представляющими типы решений, и ветвями, представляющими их отношения.

Зависимости: Взаимоотношения между логическими заключениями и теми предпосылками вывода, на основе которых они получаются.

Зашумленные данные: Данные, ряд характеристик которых вносят неопределенность в процесс проведения рассуждений в системах искусственного интеллекта.

Знания: Накопленная информация, предназначенная для решения задач искусственного интеллекта. В случае экспертных систем знания подразделяют на факты, убеждения или эвристические правила. Для речевых систем или систем распознавания образов знания рассматривают как хранящиеся в памяти образцы, изображения, словесные данные, факты и эвристические правила.

Инженерия знаний: Дисциплина, связанная с приобретением и выделением знаний при помощи эксперта, и последующим объединением с базой знаний в экспертной системе.

Исчисление предикатов: Формальный язык символьных структур, используемый для представления фактов.

Лексикон: Список, состоящий из терминов, слов и морфем конкретного языка, обычно содержащихся в базах знаний систем понимания речи или понимания естественного языка.

Лисп (Lisp — LISt Programming): Программирование списков является языком программирования искусственного интеллекта, разработанным Дж. Маккарти в компании MIT в 1958 г.

Лисп-машины: архитектуры вычислений, предназначенные для выполнения программ на языке Лисп.

Логика высказываний: Логика, которая посредством высказываний определяет истинность или ложность других высказываний.

Логика предикатов: Логические операции, которые на основе манипуляций с предположениями позволяют сделать утверждения.

Логические выводы, ориентированные на цель: Тип логических выводов, используемых в процедурах вывода «от цели к фактам». Основное внимание уделяется изучению состояния и правил логических выводов, которые вырабатывают желаемую цель. (См. также «Вывод „от цели к фактам”» и «Рассуждения, управляемые математическим ожиданием».)

Логический вывод: Процесс перехода от предположения, истинность которого установлена, к другому предположению, истинность которого, как считается, следует из истинности предшествующего предположения.

Непрерывная речь: Обычно произносимые языковые конструкции, состоящие из потока слов из заданного словарного запаса и подчиняющиеся конкретным грамматическим правилам. Следует отличать от распознавания отдельных слов.

Нисходящая схема решения: Подход к решению задачи, в котором перемещаются от некоторого текущего состояния к начальным условиям. (См. также «Вывод „от цели к фактам”» и «Логические выводы, ориентированные на цель».)

Машина логических выводов: Часть экспертной системы, содержащая процедуры и методики, которые будут использоваться для решения задач.

Машинный перевод: Намеренное применение систем обработки естественного языка для перевода документов с одного языка на другой.

Мета: Приставка, используемая с субъектами искусственного интеллекта для обозначения имеющихся знаний о базовом слове и субъекте. (Например, под метазнаниями подразумевают знания о базе знаний и системе.)

Множество правил: Набор правил, составляющих модуль эвристических знаний.

Морфемы: Основные лингвистические единицы, имеющие значение.

Морфологический анализ: Методика, используемая для обработки естественного языка. Значение и использование слова определяется путем разделения слова на части и выяснением значения каждой части.

Обучение: Процесс улучшения характеристик системы искусственного интеллекта с помощью предшествующего опыта, на-

правленный на изменение накопленных знаний или стратегий решения задачи.

Отсечение ветвей: Действия по исключению решения или группы решений из дерева решений задачи.

Планирование: Определение порядка выполнения процессов в программах искусственного интеллекта.

Поддержание истинности: Действия по поддержанию истинности и непротиворечивости элементов базы знаний.

Подсистема дачи разъяснений: Часть экспертной системы, логически обосновывающая или объясняющая свои выводы посредством выдачи краткого изложения правил вывода и данных, использовавшихся для того, чтобы получить это заключение.

Подготовка документа: Предполагаемое применение системы обработки естественного языка. Система действует как опытный редактор, проверяя ошибки в произношении и грамматике, а также предлагая способы перефразирования текста.

Поиск в глубину: Стратегия поиска в системах, основанных на знаниях, в которых возможные решения задачи представляются в виде дерева с вершинами и ветвями. При поиске в глубину ветвь и связанные с ней решения рассматриваются для всех вершин, прежде чем передвинуться на следующую ветвь. В этом способе глубина предполагаемого класса решений может быть изучена до передвижения на следующую ветвь. (См. также «поиск в ширину».)

Поиск в ширину: Стратегия поиска, применяемая в системах, основанных на знаниях, где возможные решения задачи представляются в виде дерева, состоящего из вершин и ветвей. При поиске в ширину исследуются все ветви дерева, соответствующие определенной вершине (или уровню), прежде чем переходят к следующему уровню. Таким образом, при поиске упор делается на ширину пространства состояний. (См. также «поиск в глубину».)

Понимание документа: Предполагаемое применение системы обработки естественного языка, в которой документ прочитывается, а его содержание вводится в систему.

Правила продукции: Состоящие из двух частей утверждения, определяющие определенное действие, которое надо предпринять, когда выполняется условие предпосылки. Обычно имеют вид правил «Если... тогда...». В правилах продукции содержатся процедурные знания.

Правило: Утверждение «Если... тогда...», используемое при проведении дедуктивных рассуждений.

Правильно построенная формула: Синтаксически правильное утверждение в исчислении предикатов.

Прагматические знания: Знания о высказываниях людей, относящиеся к общему контексту, по которому записываются или произносятся слова и фразы, а также знания о том, как различные фразы связаны друг с другом.

Приобретение знаний: Выделение и формулирование знаний для использования в системах искусственного интеллекта.

Программная среда экспертной системы: Виды представления знаний и механизмы проведения рассуждений экспертной системы, не включая базу знаний по конкретной области.

Программные средства знания: Программные средства, позволяющие эксперту, мало знакомому с инженерией знаний, запрограммировать знания в систему искусственного интеллекта.

Пролог (Prolog — Programming in LOGic): ПРОграммирование в ЛОГике, язык программирования, ориентированный на логику, был разработан Л. Колмерауэром и Р. Росселом в Университете г. Марсея в 1973 г.

Пространство решений: Концептуальный способ обдумывания возможных решений задачи, прямо определяющий число ветвей, которое система искусственного интеллекта может искать при решении задачи.

Процедурные знания: Знания о действиях и процедурах, обычно определяемых правилами продукции «Если... тогда».

Прямой логический вывод: Тип логических выводов, используемый при прямом построении цепочки логических рассуждений «от фактов к цели». При этом логический результат получают, применяя правила вывода к имеющимся данным или условиям.

Рассуждения, управляемые математическим ожиданием: Процедура управления, использующая математическое ожидание для формулировки гипотез о ненаблюдаемых ситуациях. (См. также «Вывод „от цели к фактам”» и «Прямой логический вывод».)

Рабочая область общего доступа: Система архитектуры, используемая во многих системах искусственного интеллекта, в которых промежуточные или частичные результаты деятельности по решению задач записываются и распределяются по более чем одной базе знаний.

Распознавание речи: Распознавание устной речи людей компьютерной системой.

Речь, произносимая раздельно: Речь, имеющая паузы между словами.

Семантика: Небуквенная интерпретация значений слова. Знания в данной области широко используются при интерпретации речи и естественного языка.

Синтаксис: Использование слов для образования фраз, частей предложений и предложений.

Семантическая сеть: Схема представления связей между объектами в базах знаний и системах искусственного интеллекта, использующая понятие классов эквивалентности и наследования знаний.

Синтаксический анализ: Процесс разбиения предложения или фразы на их составные фразы с целью идентификации формы, функции и синтаксических связей для каждой из частей.

Системы распознавания речи, зависящие от диктора: Системы распознавания речи, которые могут принимать входной сигнал от конкретного диктора.

Системы распознавания речи, не зависящие от диктора: Системы распознавания, которые могут понимать речь любого диктора.

Слот: Одиночное описание объекта во фрейме. Слоты могут содержать такую информацию, как название, цвет, определение или величину.

Сопоставление с образцом: Один из способов, которым идентифицируют изображения и произнесенные слова. Матрицы, которые обрабатывают зависящие от времени частотные сигналы (звук для слов, свет для изображений) и сравнивают полученные сигналы на предмет сходства с известными изображениями и словами.

Список свойств: Конструкция в программе на языке Лисп, связывающая свойства и соответствующие значения с каждым атомом программы. Списки свойств описывают состояние «окружающего мира» в программах искусственного интеллекта и, следовательно, часто обновляются.

Сценарий: Структура данных, представляющая последовательность событий. Сценарии представляют процедурные знания и в принципе весьма близки фреймам.

Техническое зрение: Понимание и интерпретация сцен или образов компьютерной системой.

Убеждение: Гипотеза об исходе каких-либо ненаблюдаемых или неопределенных ситуаций, обычно вырабатываемая машиной в процессе решения задачи.

Управление: Определение общего порядка или организации процедуры и действий по решению задачи.

Утверждение: Утвердительно высказывание или заявление.

Факт: Небольшой фрагмент декларативных знаний.

Фрейм: Структуры данных в системах искусственного интеллекта, представляющие объекты в виде списков признаков, их свойств и отношений к другим объектам, которые могут быть обработаны как единичный объект в базе знаний.

Шкала полутонов: Символьные или цифровые показатели, соответствующие полутонам серого цвета. Максимальные и минимальные числа представляют белый и серый цвета.

Эвристический поиск: Использование знаний о конкретной задаче или области для снижения или сокращения общего размера пространства поиска.

Экспертиза: Возможность воспользоваться высококвалифицированными знаниями, навыками и опытом в конкретной области. В контексте систем искусственного интеллекта сюда включают средства, допускающие интеллектуальные операции (такие как метазнания), процедуры эвристического поиска, правила выполнения логических выводов.

Экспертная система: Компьютерная система, дающая возможность воспользоваться высококвалифицированными знаниями и опытом для решения таких задач, где для людей потребовались бы годы специального образования и практической подготовки.

Языки программирования, ориентированные на логику: Языки программирования, цель которых заключается в программировании и решении логических задач.

Языки программирования, ориентированные на объект: Языки программирования, целью которых является описание характеристик объектов и связей между ними.

ОПТИЧЕСКИЕ МЕЖЭЛЕМЕНТНЫЕ СОЕДИНЕНИЯ ДЛЯ СИМВОЛЬНОЙ И ЦИФРОВОЙ ОБРАБОТКИ И ЛОГИЧЕСКИЕ ЭЛЕМЕНТЫ

*А. Д. Мак-Аулэй, Отделение вычислительной науки
и техники, Государственный университет им. Райта,
Дайтон, шт. Огайо*

11.1. Введение

В настоящее время в области высокоскоростной обработки информации наблюдается смещение интереса в раздел символьных и параллельных вычислений. Символьные вычисления позволяют использовать методы искусственного интеллекта (ИИ), а параллелизм обеспечивает увеличение скорости обработки, необходимое для удовлетворения все возрастающих требований к вычислениям. Процессоры ориентированы на выполнение одной-единственной операции, после чего передают результат на следующий процессор. Это позволяет добиться высокого уровня параллельности при обработке данных. В свою очередь оптические методы способны обеспечить высокий уровень параллелизма и высокую степень широкополосности, необходимые для реализации связи процессоров в таких системах. В гл. 11 описана система, состоящая из обрабатывающих элементов и оптического коммутатора, построенного на основе ПМС. Это коммутирующее устройство обеспечивает реконфигурацию и поддерживает концепцию потокового программирования. Обсуждаются преимущества рассматриваемой схемы относительно других подходов.

На базе процессора, описанного в данной главе, можно реализовать экспертные системы ИИ (основанные на правилах), коррелятор и матрично-векторный умножитель. Назначение процессоров и установок и переключений обеспечивают эффективность выполнения символьных и численных вычислений. Авторегрессивная (АР) процедура моделирования, применяемая в задачах спектрального анализа, уплотнения данных или оптимальной фильтрации, приведена в целях демонстрации возможности эффективной реализации сложных алгоритмов.

Символьно-цифровой распознаватель речи, выполненный на основе предложенного процессора, обладает возможностью одновременно распознавать и символы, и числа. Ранее упомянутый коррелятор и АР-моделирующее устройство используются для определения коэффициентов линейного прогнозирования исходов (ЛПИ), описывающих речевые фреймы. С помощью ди-

намического изменения временного масштаба устанавливается корреляция между полученными коэффициентами ЛПИ и коэффициентами ЛПИ единиц словаря фреймов, описывающих выделяемые слова. Система, основанная на правилах, анализирует флексии слов и выполняет распознавание. Основанный на правилах синтаксический анализатор повышает точность распознавания путем использования грамматических правил для нахождения аналогично звучащих слов из большого словарного запаса.

Из анализа современных тенденций следует, что вычисления в реальном времени будут включать в себя символьную обработку и характеризоваться высокой степенью параллелизма. В этом контексте оптика может обнаружить преимущества над традиционными полупроводниковыми технологиями в целом ряде применений, потому что она способна обеспечить очень высокую степень параллелизма, большую ширину полосы частот между процессорными связями, а также высокие коэффициенты объединения по входу и разветвления по выходу. Очень часто обработка данных в реальном времени требует быстрого принятия логических решений, использующих накопленные знания и/или обработку больших объемов данных с высокими скоростями. Таким образом, тесная связь между символьными и цифровыми вычислениями достаточно часто является полезной и желательной. Об этом свидетельствуют многочисленные примеры из таких областей, как понимание речи и изображений, робототехника, военная техника, управление промышленным производством. В статье описаны символьные и цифровые процессоры обработки сигналов с оптической перекрестной схемой межэлементных соединений. В качестве примера описаны такие системы, основанные на правилах, как коррелятор и умножитель матрицы на вектор (реализуемые на основе ЭВМ). Авторегрессивное моделирование показывает возможность выполнения более сложных алгоритмов. Описаны устройства распознавания речи и синтаксического анализа, демонстрирующие то, как с помощью предложенного оптического процессора могут быть успешно совмещены символьные и цифровые вычисления.

11.2. Роль параллелизма и символьной обработки в вычислениях, проводимых в реальном времени

11.2.1. Необходимость символьных вычислений

Широкое использование компьютеров дома и в офисах и развитие таких отраслей знаний, как искусственный интеллект (ИИ) и робототехника, привлекли внимание к тому факту, что в настоящее время возрастает число и объем задач, требующих символьных вычислений. Например, обработка слов включает

поиск и преобразование символов. Фактически многие программные средства, применяемые в компьютерах, такие как редакторы, компиляторы и отладчики, широко используют символьную обработку. Приключенческие игры, такие как «Подземелье» или «Драконы», также используют символьную обработку.

На основе символьных вычислений появляются новые методы решения задач, выходящие за рамки цифровых и статистических подходов, поскольку качественная информация или априорные знания могут быть представлены в доступной для обработки форме, а именно в виде баз данных или процедур. В частности, важной особенностью метода оптических межэлементных соединений, представленного в данной главе, является способность реализовывать несколько интерфейсов между символьными и цифровыми вычислениями.

11.2.2. Необходимость вычислений в реальном времени

Потребность решать практические задачи в реальном времени, подобно человеку (например, в робототехнике, речевых системах, системах технического зрения, системах управления автономных транспортных средств), наталкивается на необходимость выполнения огромных объемов символьных и цифровых вычислений. Это обусловлено большими объемами необходимой априорной информации и высокими скоростями поступления данных от датчиков. В будущем символьные машины, несомненно, станут более распространенными, это произойдет тогда, когда они будут способны выполнять более широкий круг задач, чем в настоящее время. В свою очередь технология создания роботов, например, будет развиваться более быстро, если логические дедуктивные методы и обработку данных от датчиков можно будет выполнять в реальном времени с приемлемой ценой.

Увеличение скорости вычислений позволит выполнять в реальном времени многие операции, ранее выполнявшиеся автономно. Возьмем пример из геофизики, где корабль или экипаж непрерывно собирает данные по нефтеразведке. Зачастую конкретные выводы по полученным данным удается сделать со значительным опозданием, когда экипаж уже находится в сотнях миль в стороне. Возвращение команды на ранее разведанное место для повторных измерений обычно не допустимо (слишком высокая цена). Следовательно, приходится затрачивать усилия на попытки исправить искаженные данные. Ясно, что способность вести обработку в реальном времени позволила бы провести повторные измерения или даже изменить стратегию сбора данных в результате наблюдения за уже обработанными данными.

Цифровые вычисления для автоматизированного проектирования, изготовления и тестирования изделий часто представляют наиболее слабые места в производственном цикле. В этих случаях цель состоит в обеспечении производительности, позволяющей пользователю работать в реальном времени. Для таких вычислений могут быть рассмотрены режимы псевдореального времени.

11.2.3. Необходимость параллельной обработки

Необходимость выполнения вычислений в реальном времени (или увеличения скорости вычислений), а также все большее распространение компьютеров требуют увеличения скоростей вычислений, а соответственно и наращивания вычислительных мощностей. Существующие сегодня подходы к разработке компьютеров выходят на уровень насыщения. Технология полупроводниковых устройств достигла той точки, где становится все труднее достичь улучшения характеристик путем уменьшения размеров компонент и, следовательно, с помощью однопроцессорных систем [1, 2]. Разрабатываемые новые технологии для традиционных подходов, таких как GaAs-приборы, обеспечат расширение возможностей в течение лишь нескольких ближайших лет; ожидается, что они обеспечат улучшение характеристик приблизительно в пять раз. (Устройства на GaAs также исследуются в целях создания матриц параллельной обработки на основе оптических бистабильных элементов.) Операции с векторами, конвейерная обработка специальных функций, работа с совмещением операций и выполнение векторных команд могут достичь значительного выигрыша в характеристиках, но они уже сегодня используются в полной мере и едва ли дадут в будущем существенное улучшение характеристик.

Для удовлетворения возрастающих требований к символьным и цифровым вычислениям наиболее перспективным подходом является широкое использование параллелизма при обработке. Параллелизм уменьшает время, затрачиваемое на вычисления. Однако если задача связана с использованием устройства ввода-вывода, то более быстрые вычисления не дадут преимуществ. Чтобы воспользоваться преимуществами параллельной обработки, требуется увеличение ширины полосы частот устройств ввода-вывода за счет увеличения или числа каналов, или их быстродействия.

Параллелизм обработки также позволяет проводить более обширные вычисления с конкретными данными. Во многих областях успехи в развитии компьютерной технологии используются для увеличения вычислительных возможностей, а не для экономии средств. Пользователь имеет ограниченное количество устройств ввода-вывода, и по мере возрастания произво-

длительности компьютера пользователь будет стремиться выполнить больший объем вычислений с теми же данными, а не увеличивать число типов данных. Примером служит обратная задача в геологии [3, 4].

В данной главе основное внимание уделено символьной и цифровой обработке, потому что в таких системах обычно дается большее значение быстродействию, чем гибкости организации прогноза данных или даже эффективности использования аппаратного обеспечения ЭВМ. Таким образом, параллелизм является лишь средством достижения высокой скорости обработки данных.

К настоящему моменту параллелизм обработки является сложным и трудно реализуемым средством увеличения вычислительных возможностей, применяемых к конкретным задачам. Причина этих трудностей и способ применения оптических методов для решения этих задач обсуждаются в разд. 11.2.4 и 11.3.

11.2.4. Требования, предъявляемые к структуре процессоров

Спорным является вопрос о размере процессора в параллельной системе. Несколько соединенных между собой компьютеров Сгау представляют собой одну крайность и включают в себя миллион одноразрядных процессоров. С другой стороны, использование от 10 до 30 достаточно мощных процессоров представляет собой традиционное средство удовлетворения сегодняшних вычислительных потребностей, что обусловлено тем, что до настоящего времени отсутствуют способы достаточно простого и эффективного параллельного соединения большого числа процессоров.

Необходимы, однако, методы, линейно наращивающие свои возможности по мере значительного увеличения числа процессоров.

Между тем при параллельных вычислениях возникает проблема, связанная с тем, что существуют задачи, которые не удается реализовать на уровне распараллеливания крупных структурных блоков обрабатываемых элементов. Общее быстродействие определяется именно этими частями, а вследствие синхронизации различных процессоров снижается общая эффективность работы системы. Ситуация оказывается аналогичной случаю, когда повышению быстродействия машины, предназначенной для обработки векторов, препятствует наличие 10% скалярных данных. Как бы быстро ни осуществлялся прогон данных, относящихся к векторам, обработка скалярных данных занимает то же самое время что и ранее, и максимальный выигрыш в быстродействии составляет 10 раз. Однако, как будет показано далее, применение большого числа простейших процессоров позволяет фактически достичь параллелизма с по-

мощью методов управления потоком данных, в то время как на уровне крупных структурных элементов схемы обработки параллелизм отсутствует. Чтобы сделать такой «параллелизм» отчетливо различимым для пользователя, необходимы изощренные программные средства.

По экономическим причинам при приемлемой цене и выбранном уровне «зернистости» структуры необходимо использовать процессоры с наибольшим быстродействием. Большое число быстрых простейших процессоров требует, однако, сложной и широкополосной сети соединений между процессорами. Для цифровых и символьных вычислений оптика представляется единственной технологией, обладающей способностью обеспечить необходимую ширину полосы частот.

11.3. Преимущества перекрестных оптических схем

11.3.1. Проблемы параллельной обработки

Предложено много способов достижения параллелизма обработки [1] (рис. 11.1). При проработке предложенных быстродействующих устройств соответственно рассматривались возникающие при этом трудности. Распределение загрузки между рядом процессоров не снимает требования иметь хотя бы минимальный уровень централизованного управления. При этом также необходимы и соответствующие каналы связи с центральным запоминающим устройством.

Распределение ресурсов памяти между процессорами может привести к возникновению конфликтных ситуаций, когда два или более процессора запрашивают информацию из одного и того же места. Разрешение конфликтной ситуации дополнительно увеличивает сложность, цену и неэффективность выполняемых вычислений. Простой пример такой ситуации можно привести из области умножения матрицы на матрицу, где для одновременного умножения на каждый из столбцов второй матрицы должна одновременно присутствовать во всех процессорах одна и та же строка первой матрицы. В подобных ситуациях, для того чтобы избежать конфликтной ситуации с запоминающим устройством, может использоваться стратегия запрограммированного потока данных, которая предусматривает переключение потока данных.

На практике на основе имеющегося опыта требуется разбить задачу на отдельные ступени таким образом, чтобы различные процессоры завершали свои задания в подходящие моменты времени, обеспечивая информацией следующие ступени обработки. При этом система синхронизации заставляет дожидаться окончания работы наиболее медленных звеньев, вызывая неэффективную работу всего вычислительного устройства

в целом. Для того чтобы, так или иначе, избавить пользователя от этого негативного обстоятельства, требуются соответствующие программные средства. А именно, необходимы специальные алгоритмы параллельной работы, устойчивость и точность которых следует сравнить с последовательными алгоритмами.

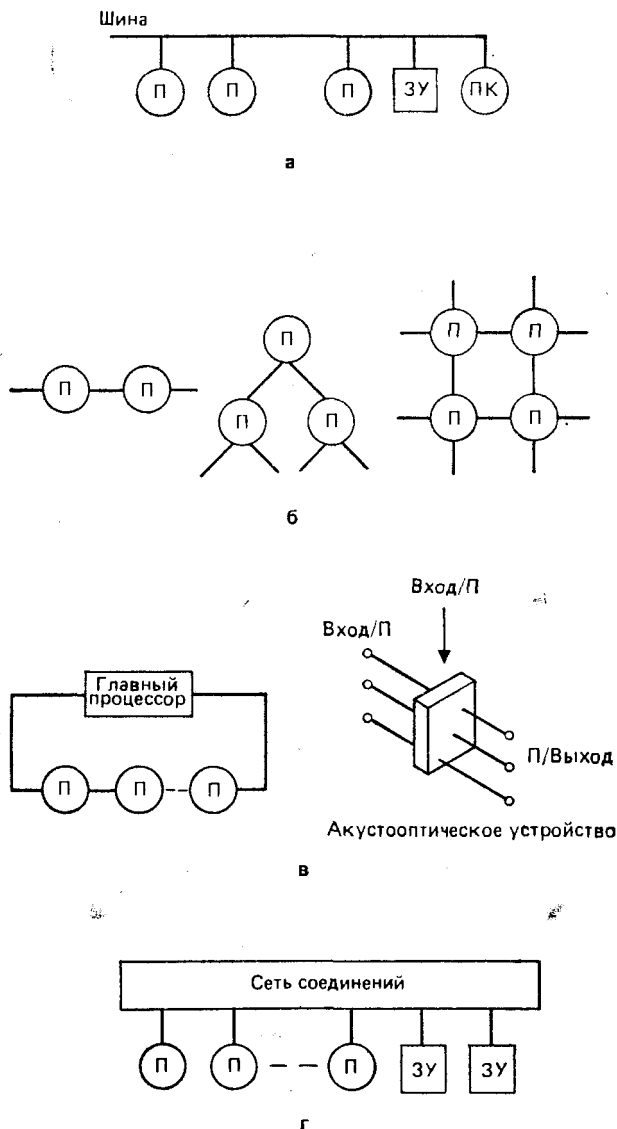


Рис. 11.1. Типы параллельных архитектур.

П — процессор; ЗУ — запоминающее устройство; ПК — переключатель

Процедура вычислений, связь с устройствами ввода-вывода, внутрипроцессорная связь — все это должно быть приведено в соответствие с имеющимися ресурсами с целью их оптимального использования; при этом существенную роль играет их относительная цена. Это выполняется с помощью тщательного подбора процессора, канала связи с соответствующей полосой частот и объема локальной памяти в каждом процессоре. В этом случае также может оказаться полезной компиляция процедур языка высокого уровня в язык низкого уровня обработки потока данных, что облегчает распределение нагрузки.

11.3.2. Преимущества оптических межэлементных соединений

Оптические волокна характеризуются большей шириной полосы частот, чем кабели вследствие уменьшенной емкостной нагрузки. Широкая полоса частот делает возможным применение последовательной передачи данных между двумя быстродействующими процессорами. Это требует меньшего числа соединений, чем потребовалось бы при использовании параллельных каналов. При высокочастотной передаче по кабелям генерируется электромагнитное излучение, вызывающее перекрестные наводки. Кроме того, электромагнитные поля, проходящие через петли, образованные соединяющими кабелями, генерируют интерферирующие сигналы. Оптические волокна, однако, обладают «иммунитетом» к этому типу интерференции волн.

Оптика может иметь преимущество с точки зрения реализации более высокой степени трехмерной интеграции. Линзы обеспечивают большие значения коэффициентов разветвления по выходу и объединения по входу, которые трудно получить в электронике. Высокая степень параллелизма достигается за счет прохождения изображений через двумерные ПМС. Достигнутый уровень оптических ПМС позволяет предполагать, что перекрестные переключатели могут стать доступными по цене, быстродействию и размерам, существенно превосходя обычные полупроводниковые устройства по характеристикам.

Высказывались предположения, что использование оптики для связи с интегральными схемами со сверхвысокой степенью интеграции позволит преодолеть такие проблемы, как ограничение ширины полосы частот на разъемных соединениях [5] и сложность соединения мультипроцессоров [6]. В настоящее время промышленность средств связи разрабатывает оптические переключающие устройства, которые позволят избежать возврата к электронным устройствам переключения света в волоконных структурах.

Оптические цифровые компьютеры с описанными в данной главе конструкциями могут быть реализованы уже через несколько лет. Другие подходы цифровых оптических вычислений

предлагались, например, для решения задачи «определения ближайшего соседа» [7], а также применительно к нейронным сетям [8].

11.3.3. Возможность реконфигурации сетей с оптическими переключателями

Жесткая связь, осуществляемая с помощью кабелей, обеспечивает наибольшее быстродействие системы для реализации как обработки определенного цифрового сигнала, так и простейшей экспертной системы, но при этом не обеспечивается требуемая гибкость при использовании различных алгоритмов. В случае использования десяти и более высокопроизводительных процессоров [9] шины, связывающие процессоры, входят в насыщение. С точки зрения соединяющих сетей важным обстоятельством является то, что для некоторых алгоритмов оказываются эффективными соединения между ближайшими процессорами, что позволяет снизить затраты на вычисления. Например, двумерный матричный процессор MPP, созданный в НАСА, оказывается эффективным для усиления контуров. Систолические матрицы [10, 11], аналогичные созданным в Университете Карнеги—Меллона [12], в компаниях ESL [13, 14], Hughes [15], NOSC [16, 17] оказываются эффективными для решения задач линейной алгебры, вычисления свертки, корреляции и ряда других операций. Указанные операции, с легкостью допускающие конвейерную обработку, согласно некоторым прогнозам, в ближайшем будущем будут преобладать даже в традиционной обработке сигналов. Однако алгоритмы, требующие более сложных соединений, выполняются менее успешно и с трудом реализуются в таких процессорах, особенно в автоматических системах. Большие быстродействующие оптические переключающие системы делают возможным необходимое перестроение процессора, что позволяет, например, получить общий вид и характеристики, как у системы с жесткими кабельными соединениями, систолической матрицы или более сложной сети. Запрограммированный поток данных снижает непроеизводительные затраты времени на вычисления адресов памяти, декодирование команд, а также выборки команд или данных из памяти.

Сети общего назначения [18] любого размера и с любым быстродействием при традиционных полупроводниковых технологиях требуют слишком больших затрат. Следовательно, существующие в настоящее время системы имеют ряд каскадов, требующих размещения между входом и выходом или нескольких переключателей, или переключателя, частично реализующего функции перекрестного переключателя. Машина Butterfly с широкополосной сетью связи [19] имеет ряд каскадов перекрестных переключателей 4×4 , соединенных с прекрасной си-

стемой перемещения содержимого памяти, а в вычислительном устройстве TRAC, используемом в системе ультразвукового контроля Austin [20], применяются переключатели 2×2 с конфигурацией «баньян» (Banуan). Многокаскадность увеличивает потери времени в системе и усложняет управление.

Прохождение сообщений еще более увеличивает непроизводительные издержки, приводя к дополнительным потерям быстродействия ради достижения большей гибкости. Система iPSC, созданная компанией Intel, основана на схеме «космический куб» ("cosmic cube") [21], имеет $2^7 = 127$ вершин и обладает распределенной памятью в вершинах 7-мерного гиперкуба. Каждый процессор соединяется с 7 другими процессорами из 128. Это обеспечивает большую гибкость алгоритма и большую степень сложности управляющих операций, чем у систолической матрицы, но дает меньшую гибкость, чем у полностью перестраиваемой системы.

Оптические сети межэлементных соединений обнаруживают много преимуществ по сравнению с электронными сетями [22, 23]. Например, описанный в гл. 11 процессор обработки сигналов способен полностью перестраиваться и использует заранее запрограммированные оптические переключатели, что позволяет получить высокое быстродействие, меньшие задержки и более простое управление, чем у многокаскадных перестраиваемых систем.

11.3.4. Получение запрограммированного потока данных с помощью оптических переключателей

Поток данных — это способ реализации направленного графа для алгоритма в подходящей для этого машине с перестраиваемой конфигурацией [24—26]. Оператор срабатывает, как только получит все необходимые для него значения. В большинстве случаев, где предусматривается использование потока данных, гибкость и способность реализовывать рекурсивные функции имеют более высокий приоритет, чем быстродействие. При этом такие системы направлены на выполнение универсальных вычислений, а не символьных и цифровых вычислений в реальном времени. Машины обычно включают в себя динамическое назначение процессоров и передачу пакетов, содержащих информацию как о необходимых операциях и маршрутах, так и о данных. Примерами современных прототипов таких машин служат: машина, созданная в MIT [27], манчестерская потоковая машина [28], японская машина Sigma-1 [29] и машина, созданная в Texas Instruments [30].

Перестраиваемые оптические переключатели позволяют использовать запрограммированные потоки данных вместе с заданным процессором. Маршрут данных и последовательность

операций в каждом процессоре предварительно вычисляются с целью минимизации непроизводительных затрат времени при счете. В следующем тактовом цикле, после того как процессор получает все требуемые входные сигналы, он будет выполнять заранее предусмотренную операцию (определенную локальным кодом). Требуется достичь максимальной производительности и минимальных задержек посредством предварительно определенных простых процедур управления и стратегий формирования потока данных.

11.4. Оптический процессор с перекрестной схемой

11.4.1. Описание системы

На рис. 11.2 изображена принципиальная схема оптического процессора с перекрестной схемой. Все имеющиеся в схеме 512 процессоров соединяются с оптическим переключателем 1024×1024 элементов с 160 МГц (или более быстродействующей) последовательной волоконно-оптической линией связи. Процессоры имеют два входных канала от перекрестного переключателя и два выходных канала к перекрестному переключателю. Первые 256 процессоров могут иметь один вход, на который подается сигнал с цифрового датчика или из основного запоминающего устройства. Остальные 256 процессоров могут иметь один выход, также соединенный с основной памятью. Разделение процессоров на две группы в данном случае оказалось выгодным, поскольку данные могут передаваться между ними в прямом и обратном направлении. Перекрестные переключатели меньшего размера вместе с переключателями, коммутирующими обмен данными, могут быть использованы для эмуляции¹⁾ перекрестного переключателя 1024×1024 для тех событий, которые оказывается трудно обработать с помощью переключателя указанного размера. Это представляет собой способ увеличения числа процессоров при сохранении линейного роста производительности системы.

Метод использования нескольких путей в памяти позволяет получить широкую полосу частот канала связи между основной группой запоминающих устройств и процессорами. Оптические переключатели управляют каналами возврата данных в память, а именно направляют их в правильно определенную группу запоминающих устройств для дальнейших вычислений с целью уменьшения усилий по управлению памятью, уменьшения сложности адресации и сохранения достаточно высокого быстродействия [23]. Перемещение данных может быть выпол-

¹⁾ В более широком смысле под эмуляцией понимают выполнение ЭВМ команд, записанных в системе команд другой ЭВМ. — *Прим. перев.*

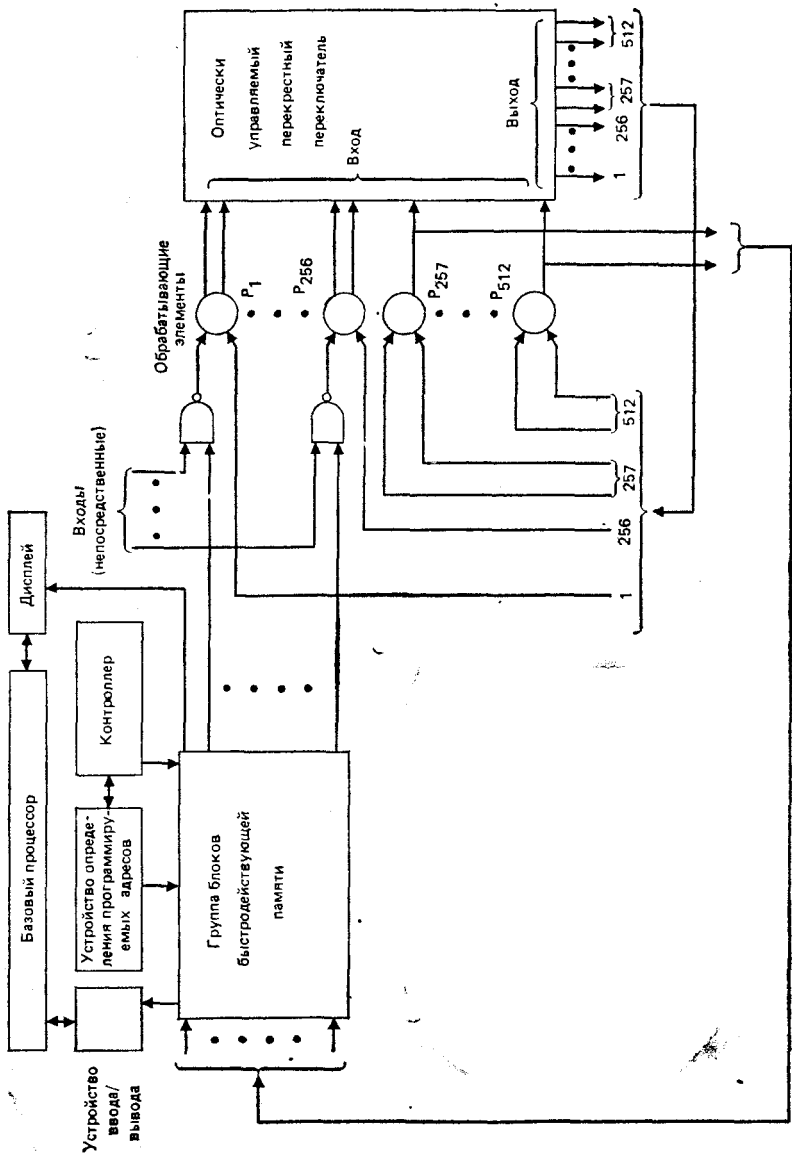


Рис. 11.2. Организация структуры оптической перекрестной обработки сигналов.

нено с помощью сброса (на несколько микросекунд) мембранного устройства на деформируемых зеркалах, осуществляемого с помощью буферного фрейма или оптическими методами. Реализация переключения оптическими методами как самостоятельный способ была продемонстрирована в [31]. В процессе обработки сигналов устройствами обсуждаемого типа вектора с большой размерностью последовательно перемещаются между памятью и процессором, при этом упрощается вычисление адресов для таких машин. Желательна реализация алгоритмов, формирующих циклическую обработку в переключателе перед обращением к памяти, что позволяет снизить требования к характеристикам и времени, затрачиваемому на перемещение данных в запоминающем устройстве [23].

11.4.2. Описание элементов обработки

На рис. 11.3 изображена структура элемента предварительной обработки. В матричном процессоре производства Texas Instruments, построенном на сверхбыстродействующих интегральных схемах, и в «систолическом чипе», разработанном в Университете Карнеги — Меллона [12], используются программируемые перекрестные переключатели. Данные, поступающие из главного оптического перекрестного переключателя и представляющие два операнда в операциях процессора, вводятся в процессор слева и преобразуются из последовательного кода в параллельный. Операнд также может поступать из основного запоминающего устройства, локальной процессорной памяти или цифрового сенсора. Сигнал, выходящий из процессора с правой стороны, вводится в параллельно-последовательный преобразователь для возврата в основной оптический перекрестный процессор. Выходной сигнал может также поступать в основное запоминающее устройство. Кроме того, могут использоваться оптические связи между процессорами и группами узлов основного запоминающего устройства.

На рис. 11.2 показано, что верхняя половина процессоров по-разному соединяется с нижней половиной процессоров. Обычными операциями процессора управляют с помощью локальной программы, а инициирование вычислений или цикла перемещения осуществляется основным синхронизирующим сигналом.

Для получения миллиарда цифровых операций с плавающей запятой в секунду необходимо использовать промышленно освоенные чипы умножителей и сумматоров с 240 нс тактовым циклом. Символьные вычисления с высоким быстродействием могут потребовать разработки специальных блоков, выполняющих сопоставление с образцом. В обычных условиях для обрабатываемых элементов не требуется наличия в них сразу и цифровых, и специальных символьных вычислительных блоков. Это

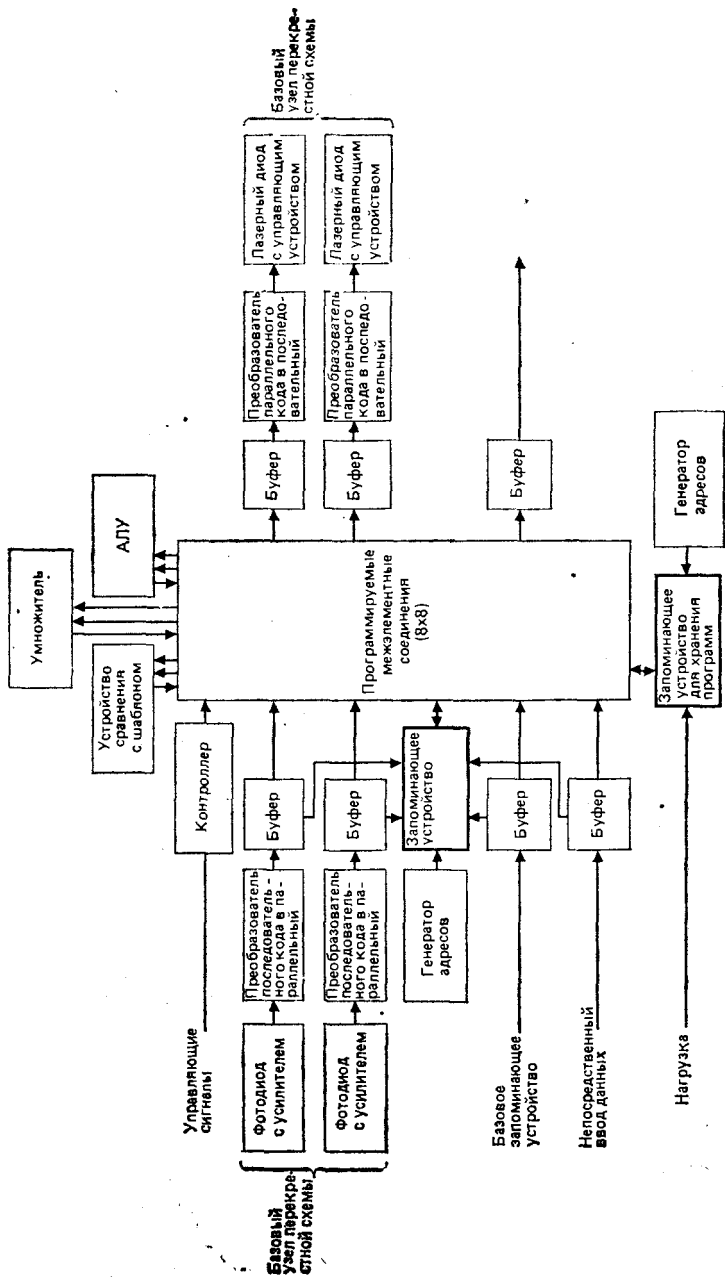


Рис. 11.3. Типичный обрабатывающий элемент.

требование скорее относится ко всей системе в целом, которая должна иметь ряд процессоров одного вида и ряд процессоров другого вида.

11.4.3. Оптический перекрестный переключатель и устройства с деформируемыми зеркалами

На рис. 11.4, а показана схема перекрестного переключателя. На каждом из пересечений горизонтальных и вертикальных линий имеется специальный переключатель, позволяющий соединять горизонтальную входную линию с вертикальной выходной. Черные кружки обозначают перекрытый переключатель. На один выход поступает информация лишь с одного входа, но с одного входа информация может передаваться на несколько выходов. На рис. 11.4, б показано схематическое изображение перекрестного переключателя, реализованного на основе ПМС. Точки обозначают области прозрачности ПМС, соответствующие установленному состоянию переключателя на рис. 11.4, а. Система линз задает горизонтальную развертку света, поступающего от входных источников, ограничивая его распространение по вертикали. Свет, проходящий через ПМС, фокусируется на фотоприемниках с помощью системы линз, собирающих свет в вертикальной плоскости, ограничивая его распространение по горизонтали.

На рис. 11.5 изображена оптическая система, основанная на применении деформируемых зеркальных устройств (ДЗЕУ), действующих как матрица перестраиваемых отражающих свет

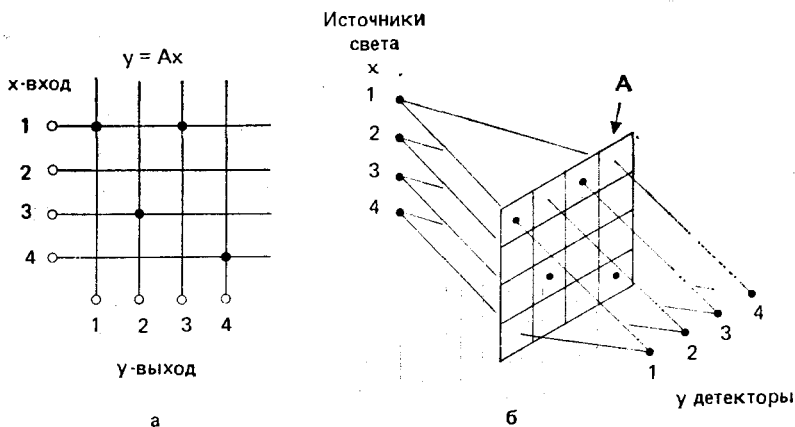


Рис. 11.4. Схематическое изображение оптического перекрестного переключателя:
 а — состояние переключателя; б — перекрестный переключатель на основе пространственного модулятора света.

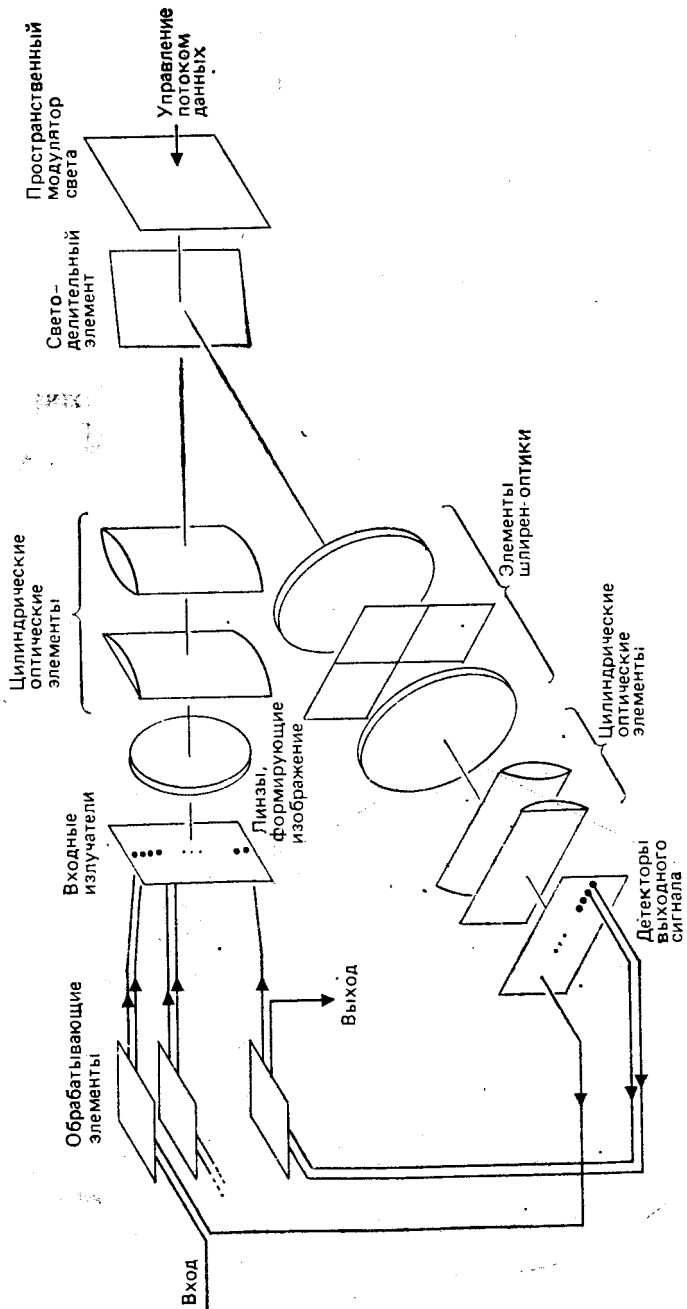


Рис. 11.5. Оптическая система для перекрестного ДЗЕУ-переключателя.

элементов, а не как модулятор пропускания. Соответственно рабочая поверхность отражающего модулятора свет не пропускает, а отражает его в обратном направлении. Светоделительный элемент используется для разделения света, подаваемого от лазера, и отраженного обратно света. Система устройств шпирен-оптики используется для блокировки света, отраженного от областей между зеркальными дефлекторными ячейками. В качестве источника света можно использовать полупроводниковые лазеры [32], которые удается модулировать с частотой до 3 ГГц. В качестве фотоприемников целесообразно использовать *p-i-n* фотодиоды [5, 6]. По мнению автора, в будущем источники и приемники света будут выполняться интегрально, непосредственно на чипах электронных схем.

Были разработаны ДЗЕУ мембранного и рычажного типа [33]. На рис. 11.6 показан фрагмент ДЗЕУ, в котором упругая мембрана покрывает сверху двумерную матрицу МДП-транзисторов, изготовленных на кремниевом чипе. Подача сигнала на транзистор вызывает локальное притяжение мембраны и приводит к появлению углубления, возникающего непосредственно над транзистором, что вызывает фазовую задержку отраженного света. Пространственный фильтр преобразует фазу света в амплитуду сигнала, отображающую состояние ячейки. Состояние матрицы транзисторов устанавливается построчно (одновременно для всех ячеек в строке) путем приложения напряжений на клеммы к верхней части устройства. Устройства с рычажной

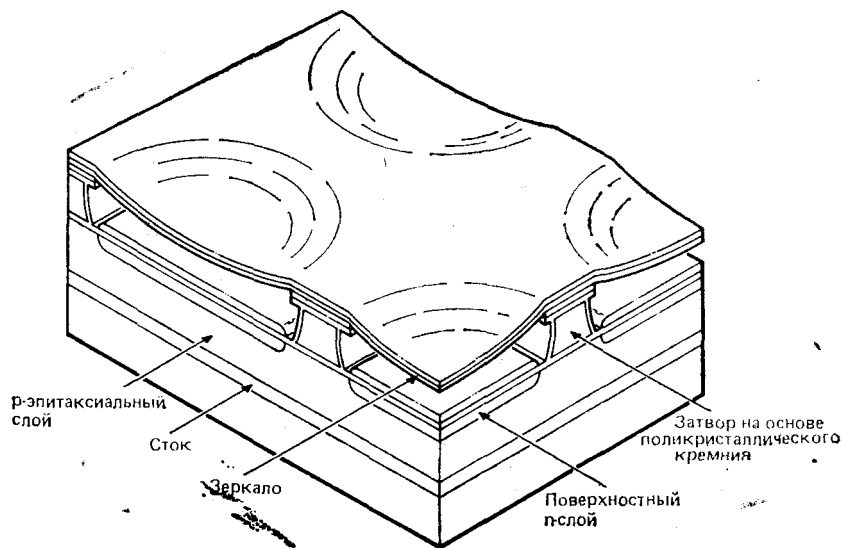


Рис. 11.6. Устройство на основе деформируемого зеркала,

системой отклонения луча имеют маленькие отражающие свет элементы, закрепленные за один угол на рычажках. Характеристики мембранного ДЗЕУ, изготовленного на Texas Instruments, и результаты выполнения с его помощью спектрального анализа описаны в [34].

Оптические перекрестные системы в целом характеризуются большей производительностью, чем электронные переключатели, но при этом одновременно и большими временами переключения. Для обсуждаемых здесь задач это не выглядит помехой, потому что все остальные элементы процессора должны успевать перестраиваться за то же время, что и оптические переключатели. На затраты времени также влияет объем данных, которые необходимо передать в сам оптический переключатель для переключения его в нужное состояние. Преимущество использования кремниевых чипов для оптических переключателей состоит в том, что схема адресации в этом случае может размещаться непосредственно на чипе. Соответственно может быть использована декодирующая схема, если, конечно, требуется переключать не слишком большое число ячеек.

11.4.4. Программирование и функциональные операции

В процессе решения задачи пользователь определяет приближенные уравнения для вычисления представляющей интерес функции и выбирает алгоритм для реализации уравнений. Направленный граф [35] составляется для алгоритма таким образом, чтобы обеспечить максимум параллелизма. Операции представляются в виде вершин, а соединения — как ребра или дуги. Направленный граф отображается на оптическую систему перекрестных переключателей: ребра — на состояния перекрестной системы, вершины — на доступные функциональные элементы. Это делается таким образом, чтобы удовлетворить временным ограничениям, и при этом эффективно использовать доступные ресурсы. Варианты потокового графа, предназначенные для реализации ряда цифровых и символьных вычислений с помощью предложенного процессора, обсуждаются в разд. 11.5, 11.6 и 11.7. Модифицированный вариант графа алгоритма вычислений «сопряженных градиентов» позволяет повысить эффективность работы обсуждаемого в данной главе процессора более 90% [36]. В настоящее время одной из наиболее важных задач является разработка программных средств для автоматического конструирования направленного графа с максимальной степенью параллелизма реализации алгоритма [37] и затем для успешной реализации его в конкретном процессоре. Это влечет за собой необходимость табличного описания состояний, включающих сети межсоединений, операций, которые должны

выполнять функциональные элементы, а также временного графика выполнения этих операций.

Пользователь вводит в процессор поток или потоки данных, и если «конвейер» был загружен, то результат появляется на выходах в каждом тактовом цикле. Каждый функциональный элемент будет при этом выполнять свою операцию за время следующего синхронизирующего импульса, а именно после того, как получит все необходимые логические или цифровые входные сигналы. При этом он будет автоматически передавать выходные сигналы на перекрестный переключатель для указания маршрута для следующей заданной операции.

11.5. Получение логических выводов с помощью оптического процессора с перекрестной схемой

11.5.1. Основные свойства экспертных систем, основанных на правилах

Простым и имеющим широкое применение видом экспертной системы является система, в которой информация содержится в форме правил продукции «если..., то...» [38]. Существует много практических задач, требующих обработки в реальном времени, в которых необходимо быстрее получение логических выводов и для которых приемлем предложенный ранее процессор. Область применения включает системы распознавания речи, технического зрения, системы управления промышленными предприятиями, робототехнику и систему вооружений. В основанных на правилах системах, построенных по схеме «от фактов к цели», проводится ряд наблюдений, и в каждом случае присваивается значение вероятности их правильности. Машина логических выводов быстро определяет состояние задачи, ближайшее окружение либо угрожающие ситуации и указывает подходящие для данной ситуации действия. В системах, основанных на правилах и построенных по схеме «от цели к фактам», цель представляется в виде гипотезы и на ее основе формулируют требования, направленные на достижение этой цели. Эти требования затем подвергаются дальнейшему разбиению на более мелкие элементы. (Подобным образом функционирует компьютерный язык Пролог.) В разд. 11.5 обсуждаются схемы «от фактов к цели» и от «цели к фактам». В обоих видах схем параллелизм необходим для достижения высокого быстродействия. Из представленных в разд. 11.5 данных следует, что системы с жесткими кабельными соединениями, использующими множество связанных между собой логических правил и процедуру присвоения вероятностей, были бы неэкономичными, поскольку в различных задачах правила могут полностью изменяться или могут нуждаться в частичных изменениях в новых

ситуациях. Для предложенного оптического перекрестного переключателя требуются высокие быстродействие и способность перестраиваться. Рассматриваемые здесь простые системы предполагают использование исчисления высказываний и не обращаются к более сложному исчислению предикатов первого порядка, что необходимо в более сложных практических задачах.

11.5.2. Схема «от фактов к цели» и ее реализация

Процессор должен непрерывно выполнять логические выводы на основе непрерывно поступающих с большой скоростью данных, что наиболее ярко проявляется в таких областях, как системы вооружений и робототехника. В таких случаях данные могут подвергаться конвейерной обработке, при этом их пропускают через оптическую систему перекрестных соединений с целью обеспечения высокой эффективности использования процессора. Для иллюстрации этого принципа используется приведенный ниже пример из книги [38].

На рис. 11.7 изображен граф направленного потока данных, служащий для идентификации семи видов животных на основе 20 наблюдаемых признаков. Упрощенный вариант получится, если предположить, что любой признак можно описать либо как ИСТИНА, либо как ЛОЖЬ и не присваивать состояниям вероятностных значений. В этом случае кружочки будут представлять вентили И, а кружочки с черными точками — вентили ИЛИ. «I» является вентилем инверсии. Выходной сигнал будет определяться как ИСТИНА, если входные сигналы, соответствующие данному животному, определяются как ИСТИНА.

Правила «если..., то», характерные для систем, основанных на правилах, изображены, например, на процессоре 3: «Если животное является млекопитающим и ест мясо, то оно является плотоядным». Вентиль ИЛИ в процессоре 21 указывает альтернативный вариант получения логического вывода о принадлежности к плотоядным: «Если животное является млекопитающим и имеет острые зубы, имеет когти и имеет направленные вперед глаза, то оно является плотоядным».

В представляющем интерес для практики случае наблюдаемые признаки характеризуются определенной вероятностью правильности.

Если предположить, что события независимы друг от друга, то общая вероятность для двух событий описывается соответствующим произведением. Следовательно, процессоры, обозначенные пустыми кружочками и не имеющие черных точек, умножают две вводимые вероятности. Ниже эти процессоры относят к типу А. Следует заметить, что выходной сигнал 0 соответствует логическому ЛОЖЬ и возникает на выходе, если любая из входных вероятностей равняется 0. Кружочки с черными точками,

Наблюдаемые
входные
признаки

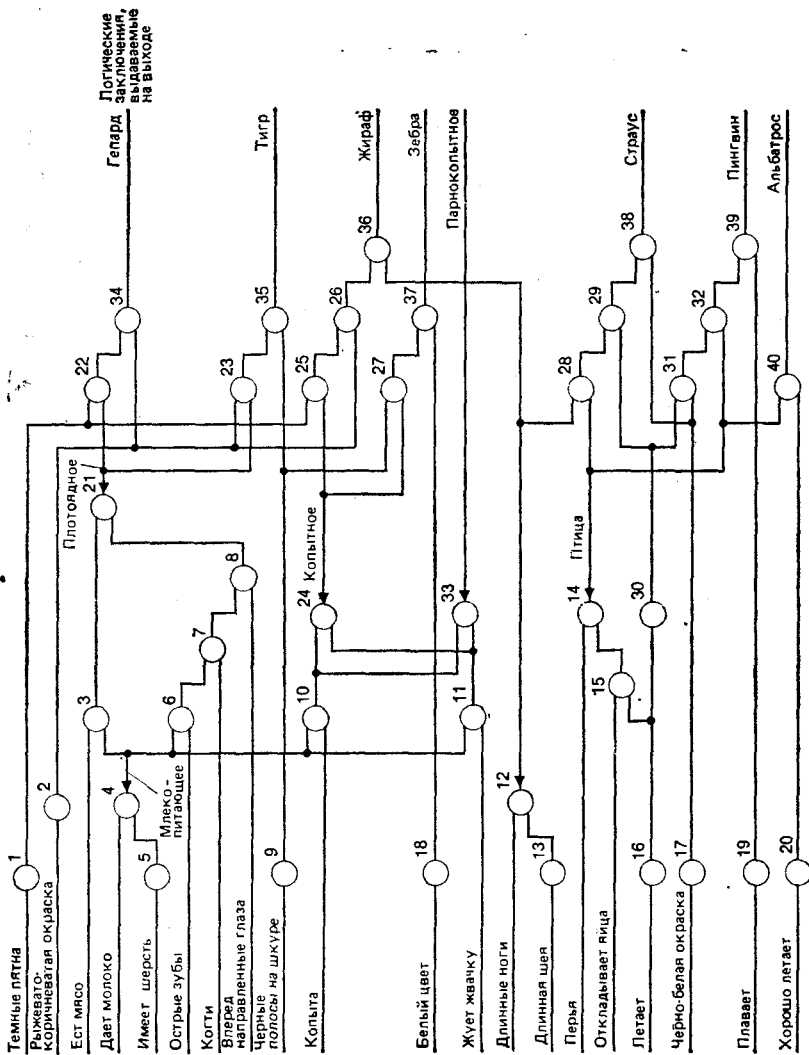


Рис. 11.7. Изображение основанного на правилах графа «от фактов к цели».

относимые ниже к типу В, определяют максимальные значения входных вероятностей. Альтернативное уравнение представлено в [38]. Процессоры типа С вычитают вводимую вероятность из 1. Все три типа процессоров также имеют функцию отображения вероятности, влияющую на выходной сигнал [38]. Функция отображает вычисленную входную вероятность на выходную вероятность, также имеющую значения от 0 до 1.

Потоковый граф на рис. 11.7 может быть реализован непосредственно на основе предложенной на рис. 11.2 системы, где верхняя половина процессоров имеет входные соединения, а нижняя половина имеет выходные соединения. Для приведенного простого примера требуется только лишь 40 процессоров. Программирование процессоров осуществляется в зависимости от того, к какому из типов (А, В или С) они относятся (табл. 11.1). Наблюдаемые входные сигналы вводятся в первые 20 процессоров. Вероятности получаются на выходах последних семи процессоров, и самое большое из значений вероятности указывает на идентифицированного животного. Получаемая вероятность оценивается таким параметром, как фактор уверенности, непосредственно характеризующим различия вероятностей для разных животных. В табл. 11.1 показано, какой из перекрестных переключающих элементов должен быть включен для обеспечения соединений между правилами на потоковом графе, приведенном на рис. 11.7.

11.5.3. Схема «от цели к фактам» и ее реализация

Многие практические задачи больше приспособлены для схем «от цели к фактам», чем для схем «от фактов к цели», поскольку пользователь заинтересован в разрешении конкретной гипотезы или цели. Получение данных может быть достаточно затруднено и пользователь не заинтересован направлять свои усилия на получение ненужных ему признаков. Можно задать законный вопрос экспертной системе, представленной на рис. 11.7: «Является ли животное гепардом»? Это предусматривает прочтение потокового графа справа налево. Цель считается достигнутой, если в процессоре 34 животное имеет пятнистую окраску, и в процессоре 22 животное имеет темные пятна и является плотоядным. Преимущество схемы от «цели к фактам» заключается в том, что она конкретно указывает, какие именно признаки надо наблюдать, и время не затрачивается впустую на получение ненужных данных.

Схема «от цели к фактам» требует решения так называемой обратной задачи, т. е. нахождения дерева, приводящего к представленной в виде гипотезы цели (рис. 11.8). Для ее реализации на основе оптического процессора с перекрестными соединениями требуется провести показанное на рис. 11.8 присвоение

Таблица 11.1. Состояния перекрестного переключателя для системы

Процессор		Состояние	Процессор		Состояние
№	Тип		№	Тип	
1	A	22а, 25в	11	A	24в, 33в
2	A	34в, 23а, 26в	12	Б	28а, 36в
3	A	21а	13	A	12в,
4	Б	3в, 6а, 10а, 11а	14	Б	28в, 32в, 40а
5	A	4б	15	A	14в
6	A	7а	16	A	15в, 30а
7	A	8а	17	A	31в, 38в
8	A	21в	18	A	37в
9	A	35а, 27в	19	A	39в
10	A	24а, 33а	20	A	40в

признаков. Однако передача сигналов вниз по этому дереву не позволяет получить конвейерную обработку, как в случае схемы «от фактов к цели». Следовательно, процессор используется не эффективно и для разработки эффективных подходов к реализации схемы «от цели к фактам» на этом типе процессоров требуются дальнейшие исследования.

11.6. Цифровая обработка сигналов с помощью оптического мультипроцессора с перекрестной схемой

11.6.1. Основные алгоритмы обработки сигналов

В работе [23] для указанного типа архитектуры проводились исследования основных алгоритмов обработки сигналов, таких, как систолическая фильтрация, свертка, корреляция и фурье-преобразование. Линейный фильтр определяют следующим образом:

$$c = a * b, \text{ или } c_n = \sum_{k=1}^K a_k b_{n-k}, \quad 2 \leq n \leq N + K, \quad (11.1)$$

где a_k , $k=1, \dots, K$ являются коэффициентами фильтра, b_n , $n=1, \dots, N$ являются значениями данных, $a * b$ — оператор свертки. За пределами заданного диапазона величин a_k и b_n предполагаются равными 0.

Корреляционная функция векторов \mathbf{a} и \mathbf{b} записывается как

$$c_n = \sum_{k=1}^K a_k b_{k+n}, \quad 1 - K \leq n \leq N - 1. \quad (11.2)$$

Сравнение уравнений (11.1) и (11.2) показывает, что свертка и корреляция могут вычисляться с помощью одного и того же

«от фактов к цели», основанной на правилах

Процессор			Процессор		
		Состояние			Состояние
№	Тип		№	Тип	
21	Б	22в, 23в	31	А	32а
22	А	34а	32	А	39а
23	А	35а	33	Б	(Парнокопытное)
24	В	25а, 27в	34	А	(Гепард)
25	А	26а	35	А	(Тигр)
26	А	36а	36	А	(Жираф)
27	А	37а	37	А	(Зебра)
28	А	29а	38	А	(Страус)
29	А	38а	39	А	(Пингвин)
30	В	29в, 31а	40	А	(Альбатрос)

процессора путем обращения порядка записи индексов одного из входных сигналов. Как обсуждалось в [39], корреляцию иногда удается вычислить даже быстрее, чем быстрое преобразование Фурье. Ниже описан метод удвоения, используемый при вычислении таких операций с помощью оптического процессора с перекрестной схемой (см. на рис. 11.2).

11.6.2. Коррелятор и матрично-векторный множитель

На рис. 11.9, а изображен граф для процедуры вычисления внутреннего произведения ряда векторов с образцовым вектором \mathbf{b} во временной области посредством рекурсивного удвоения. При умножении матрицы на вектор вектора могут рассматриваться как строки матрицы \mathbf{A} . В то время как один вектор умножается в процессорах с 1 по 8, результаты предшествующего умножения суммируются в процессорах с 9 по 12. Величина выходного вектора для умножения матрицы на вектор получается в каждом тактовом цикле, или шаге выполнения расчетов.

На рис. 11.9, б изображен вариант реализации процессора, схема которого представлена на рис. 11.2. В этом случае число переключателей колеблется от 16 до 24, а время задержки такого варианта равно $\log_2 8$. Для данной схемы требуется параллельный вход, а для заполнения конвейерной схемы обработки данными должна выполняться корреляция нескольких векторов. Корреляция (уравнение (11.2)) выполняется путем сдвига входного вектора на одну позицию по вертикали в каждом цикле.

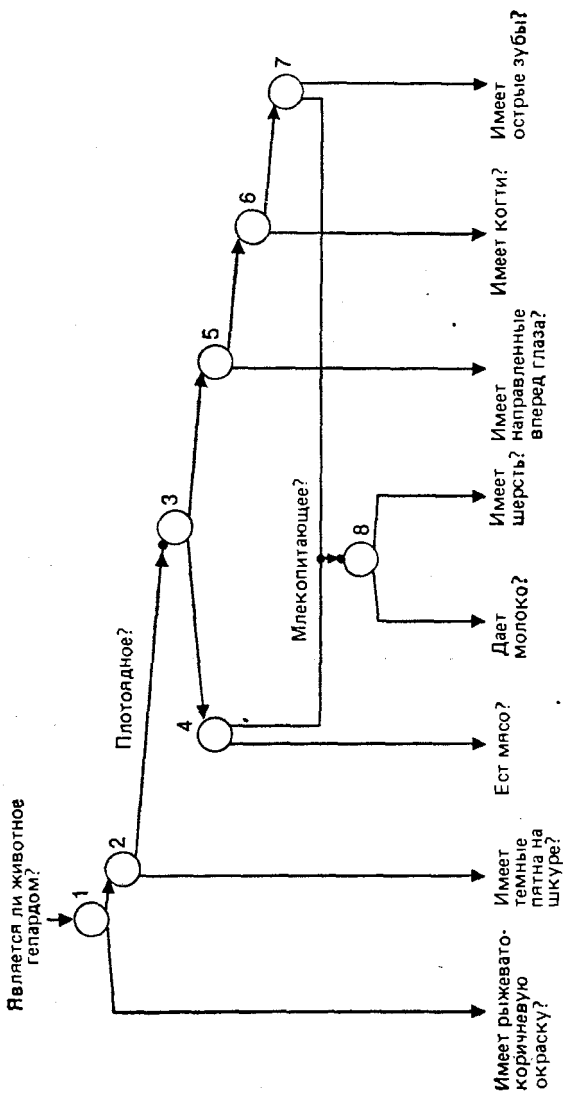


Рис. 11.8. Изображение основанного на правилах графа «от цели к фактам».

11.6.3. Авторегрессивное моделирование для задачи спектрального анализа

Авторегрессивное моделирование широко используется в обработке речи, подводной акустики, сонарах, радарх и обработке сейсмических сигналов. Авторегрессивное моделирование эквивалентно одномерному случаю линейного прогнозирования исходов, методу максимальной энтропии и является особым случаем оптимального поиска по методу наименьших квадратов или винеровской фильтрации. Цель этого моделирования заключается в представлении временных последовательностей сигналов в виде малого числа параметров авторегрессии (чисто полярная модель), с помощью которых временные последовательности сигналов могли бы быть восстановлены с точностью, определяемой методом наименьших квадратов, при пропуске белого шума через модель. Пропускание временных последовательностей через обратный фильтр удалило бы, следовательно, информацию о самих временных последовательностях сигналов и оставило бы на выходе белый шум. Поэтому параметры авторегрессии содержат информацию, необходимую для вычислений спектра временных последовательностей сигнала, т. е. его цвет. Параметры авторегрессии можно также рассматривать как коэффициенты линейного прогнозирования исходов (ЛПИ) из-за того, что конечный импульсный отклик или «пересылка среднего», рассмотренные для фильтра с этими коэффициентами, позволяют предсказать следующие значения временной последовательности сигналов исходя из предыдущих значений. Вычитание предсказанных значений из имеющихся величин дает белый шум, как и в случае применения инверсного авторегрессивного фильтра. Коэффициенты ЛПИ, а именно $a_k, k=1, \dots, m$, позволяют предсказать значение временных последовательностей в момент времени j на основании их предшествующих значений

$$\tilde{x}_j = \sum_{k=1}^m a_k x_{j-k} \quad (11.3)$$

и вычисляются при помощи минимизации выражения

$$|x_j - \tilde{x}_j|^2. \quad (11.4)$$

В связи с этим уместно упомянуть три возможных практических применения. Например, в геофизике отраженные от земли сигналы рассматриваются как случайные. Измеренные датчиками на поверхности земли сигналы представляют свертку слабых волн, испускаемых от источника, и последовательности случайных сигналов, отражаемых от земли. Влияние источника слабых сигналов исключается из данных, снимаемых с датчиков, посредством обратного преобразования свертки от прогно-

зируемых данных, т. е. в спектре сигналов исключается информация о цвете [40].

В случае обработки речи фрагмент речи, состоящий из 200 образцов, может быть представлен в виде 16-ти коэффициентов ЛПИ (параметров авторегрессии), поскольку это число полюсов приемлемо для моделирования спектра и для распознавания указанного фрагмента. В этом случае уплотнение данных размерности 200 до 16 чисел позволяет в последующих стадиях провести быстрые вычисления. Обычно для меньших словарных запасов используется 10 коэффициентов ЛПИ.

Третий пример относится к определению спектра $S(\omega)$ с помощью параметров авторегрессии a_k , $k=1, \dots, m$, и энергии v с помощью выражения

$$S(\omega) = \frac{v}{\left| 1 + \sum_{k=1}^m a_k e^{-j\omega k} \right|^2}. \quad (11.5)$$

При вычислении быстрого преобразования Фурье используют предположение о том, что за пределами измеренной области данные являются либо нулевыми, либо имеют повторяющиеся значения. В случае обработки коротких образцов данных это может приводить к закликиванию вычислений. Избежать закликивания удастся путем вычисления спектра с помощью авторегрессионных моделей.

Ранее упомянутые подходы, такие, как авторегрессия, линейное прогнозирование исходов, метод максимальной энтропии, приводят к необходимости выполнения следующих шагов по определению параметров авторегрессии или ЛПИ. Для временных последовательностей вычисляют автокорреляционную функцию

$$R_{xx}(\tau) = \frac{1}{m-\tau} \sum_{j=1}^{m-\tau} \frac{x_j x_{j+\tau}}{R_{xx}(0)}, \quad \tau=0, \dots, m-1. \quad (11.6)$$

Далее, для нахождения параметров авторегрессии или ЛПИ находят решения для \mathbf{a} , $a_k=1, \dots, m$, с помощью уравнения

$$\mathbf{R}\mathbf{a} = \mathbf{b}, \quad (11.7)$$

где \mathbf{R} — автокорреляционная матрица Топлица, образованная из R_{xx} . Эта процедура обычно выполняется на последовательном процессоре, использующем алгоритм Левинсона или Левинсона—Дарбина. Для параллельной машины [41, 62] более подходящим является алгоритм Шура. Реализация алгоритма Левинсона на основе описанного в гл. 11 процессора описана в [43], а алгоритма Левинсона—Дарбина — в [44, 42]. В данном случае используется алгоритм Шура.

Матрицу \mathbf{R} представляют в виде произведения двух «треугольных» матриц (взятых сверху и снизу от диагонали)

$$\mathbf{R} = \mathbf{U}^T \mathbf{U}. \quad (11.8)$$

Подстановка в уравнение (11.7) позволяет получить решение за два шага. Решение для \mathbf{g} получается из

$$\mathbf{b} = \mathbf{U}^T \mathbf{g} \quad (11.9)$$

и решение для \mathbf{a} находят из

$$\mathbf{g} = \mathbf{U} \mathbf{a}. \quad (11.10)$$

Автокорреляционная функция в уравнении (11.6) может быть вычислена с использованием древовидной структуры нахождения корреляционной функции, показанной на рис. 11.9, и системы, представленной на рис. 11.2. В этом случае данные вводятся в умножители (предполагают, что имеется достаточное число процессоров), и затем копия этих данных, задержанная на интервал времени, равный максимальной требуемой задержке, вводится начиная с вершины древовидной структуры. На каждом шаге корреляция выполняется с задержкой на один шаг, до тех пор пока задержанный поток данных точно не установлен по отношению к оригинальным данным. Это обеспечивает нулевое значение задержки коэффициента автокорреляции.

Коэффициенты автокорреляции вводятся в систолическую матрицу, показанную на рис. 11.10, по мере их вычисления. В этой матрице для вычисления параметров авторегрессии или ЛПИ временных последовательностей сигналов используется алгоритм Шура. Вычисляется верхняя «треугольная» матрица \mathbf{U} [41], которая затем используется в нижней систолической матрице для вычислений \mathbf{g} и \mathbf{a} . В то время как вычисляются \mathbf{g} и \mathbf{a} , два верхних столбца процессора начинают вычисление соответствующего разложения для следующей временной последовательности сигналов.

11.7. Случай одновременного выполнения символьных и цифровых вычислений, пример речевой системы

11.7.1. Преимущества одновременного выполнения символьной и числовой обработки

И обработку сигналов, и выполнение логического вывода, и связь их посредством перекрестного переключателя можно выполнять одновременно. В этом случае выходные сигналы, получаемые при обработке, могут быть непосредственно использованы в процессе получения логических выводов, которые в свою очередь запускают специальные вычисления, необходимые для завершения шага выполнения логических выводов.

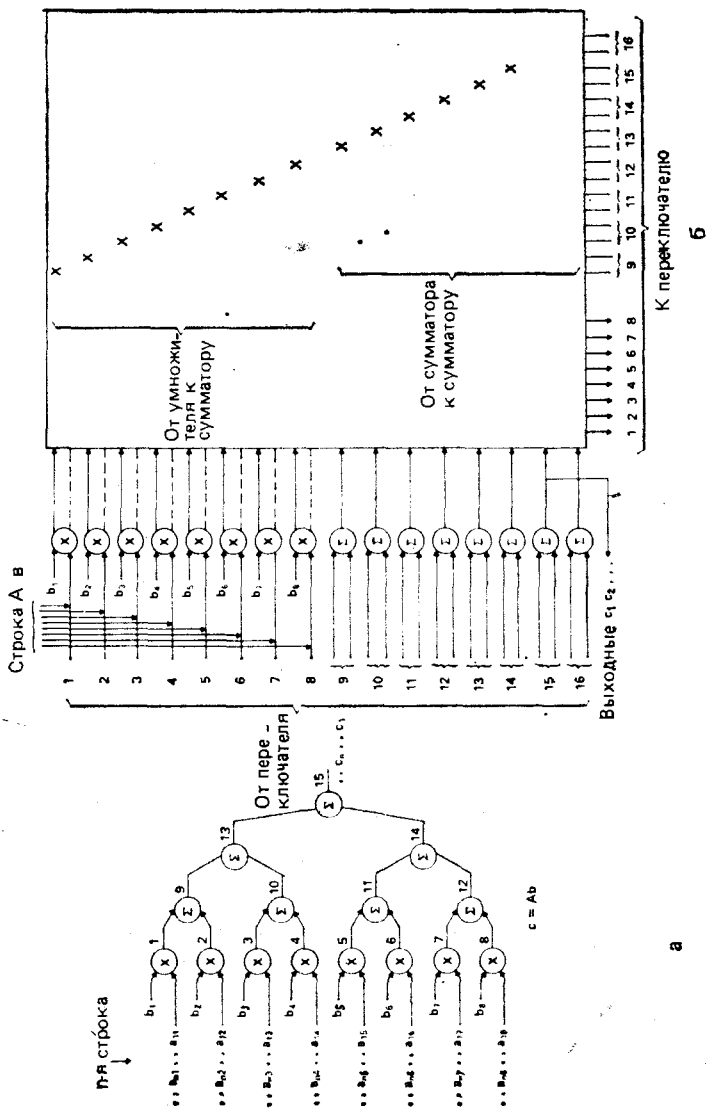


Рис. 11.9. Корреляция и умножение матрицы на вектор: **а** — граф алгоритма удвоения; **б** — реализация алгоритма удвоения на основе перекрестного переключателя 24×24 .

Практический метод реализации таких вычислений может состоять в специализации верхней части процессоров для получения логических выводов, а нижней части процессоров — на обработке сигналов. Перекрестный переключатель разделяется на четыре соответствующих сегмента, где верхний левый используется для обработки сигнала, а нижний правый для символьных вычислений. Два других сегмента позволяют осуществить связь между символьными и цифровыми вычислениями на любой стадии.

Быстрая символьная и цифровая обработка необходима в геофизических исследованиях для того, чтобы и обработку, и интерпретацию данных выполнять в реальном времени. Обработка сигналов используется для выделения собственно данных и удаления характерных особенностей источника [40]. Моделирование и решение обратной задачи [3, 4] могут тогда использоваться для оценки геофизических параметров. Выбор параметров для управления этими вычислениями и решения математических уравнений требуют использования символьных вычислений. Основанные на правилах экспертные системы предназначены для интерпретации данных и должны взаимодействовать с системой обработки сигналов. Таким образом для схем «от цели к фактам» конкретная гипотеза должна подтверждаться повторной обработкой выбранных данных.

11.7.2. Символьно-цифровое распознавание речи и система синтаксического анализа

Символьно-цифровое устройство распознавания речи иллюстрирует преимущества одновременно цифровых и символьных вычислений. На рис. 11.11 изображена блок-схема устройства распознавания речи со встроенным анализатором синтаксиса. В данном случае весьма трудная задача распознавания речи для наглядности рассматривается крайне упрощенно. Например, возможности распознавания могут быть существенно расширены по сравнению с представленными здесь с помощью метода неявного моделирования по Маркову, основанного на известных данных о порядке следования похожих слов. Произносимые данные непрерывно вводятся в схему (см. рис. 11.11) с левой стороны. Каждое слово разделяется процессором на фреймы длительностью 20 миллисекунд, по 200 образцов в каждом. Предполагаемая здесь система оперирует с непрерывным потоком фреймов в отличие от большинства систем, которые работают с блоками фреймов в пакетном режиме. Это позволяет избежать ряда сложностей, связанных с границами блоков. Два блока слева на рис. 11.11 уменьшают объем данных фрейма с 200 образцов до 16 коэффициентов линейного прогнозирования исходов с помощью АР-моделирования (см. 11.6.3). Параллель-

ный вариант реализации системы, показанной на рис. 11.2, уже был представлен ранее на рис. 11.9 и 11.10. Обычно следующие этапы, используемые для улучшения представления речи, строятся на основе эвристических подходов.

Для идентификации или классификации поступающих в систему слов (входящих в словарный запас) осуществляется динамическое изменение временного масштаба слов, что достигается путем вычисления корреляции коэффициентов ЛПИ поступающих сигналов и коэффициентов ЛПИ образцов, хранящихся в словаре. В системе, основанной на правилах, для определения начала и конца слов используется набор простых правил. В свою очередь, для устранения возникающих неоднозначных ситуаций применяется символьный анализатор синтаксиса, имеющий вид системы правил, организованных по принципу «ситуация — действие» и «жди и смотри». Указанный анализатор синтаксиса основан на грамматических правилах языка. Может быть реализован другой вид анализатора синтаксиса, выполняющего ограничение числа слов, которые следует рассматривать на следующей стадии корреляции.

Динамическое изменение временного масштаба при параллельной обработке и символьный анализатор синтаксиса описываются ниже, вместе с примером их реализации, осуществляемой с помощью оптического переключателя. За время обработки каждого фрейма перекрестный переключатель трижды изменяет свое состояние. Первое состояние относится к вычислению коэффициентов ЛПИ и получению логических выводов для анализатора синтаксиса. Второе состояние относится к нахождению

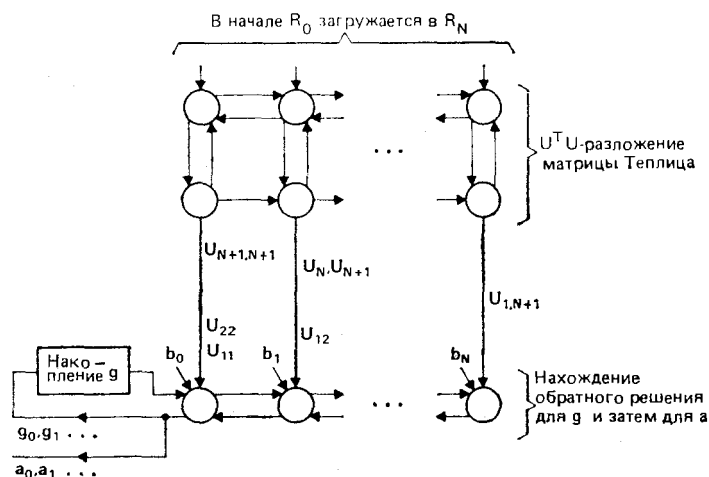


Рис. 11.10. Параллельное вычисление коэффициентов линейного прогнозирования исходов (ЛПИ) с помощью алгоритма Шура.

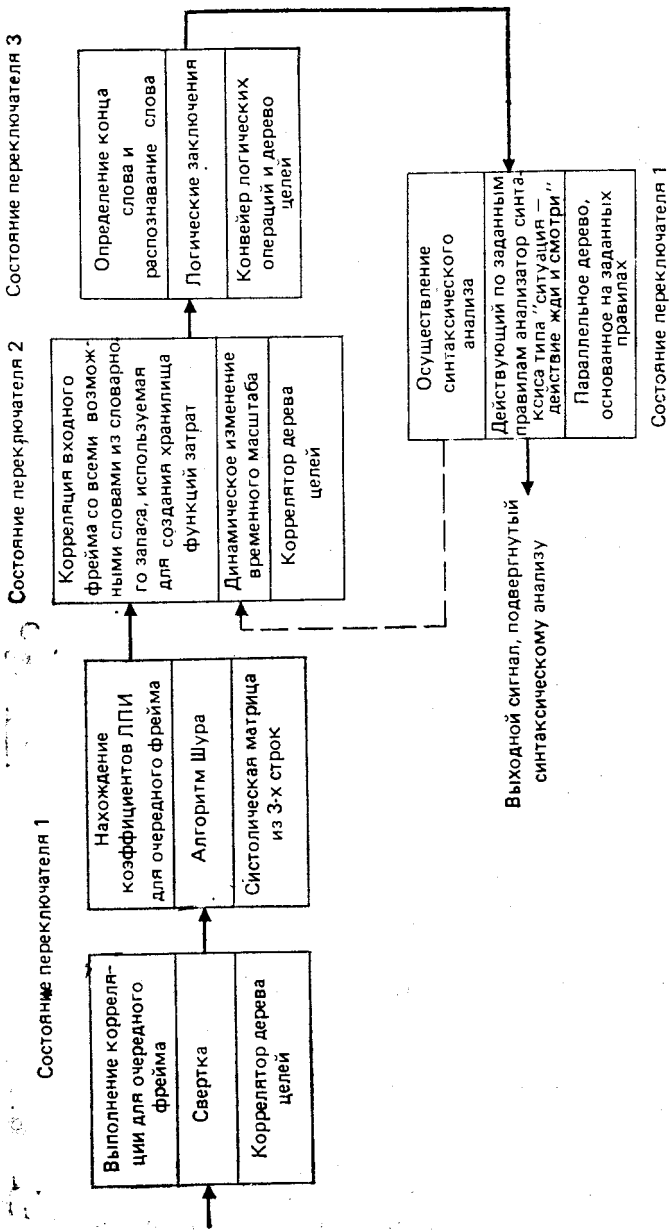


Рис. 11.11. Распознавание речи со встроенным систолическим анализатором синтаксиса.

дению функции затрат для каждого опорного слова-образца в процессе динамического изменения масштаба времени, а третье — к определению окончания каждого образца и его классификации.

11.7.3. Распознавание речи с помощью динамического изменения масштаба времени

Коэффициенты ЛПИ для фрейма входных данных должны коррелировать с каждым фреймом для любого слова, хранящегося в словаре образцов ЛПИ. На рис. 11.12 по оси X изображены входные фреймы ЛПИ, а по оси Y — фреймы ЛПИ опорных слов. Словарь опорных образцов разделен на подсловари, в соответствии с частями речи. Слова, которые могут выступать в качестве более чем одной части речи, включены в отдельные подсловари.

Типичное локальное расстояние между i -входным фреймом и v -опорным фреймом для k -го слова составляет

$$d_{i,r} = \log \sum_{n=1}^L (i_n * r_n), \quad (11.11)$$

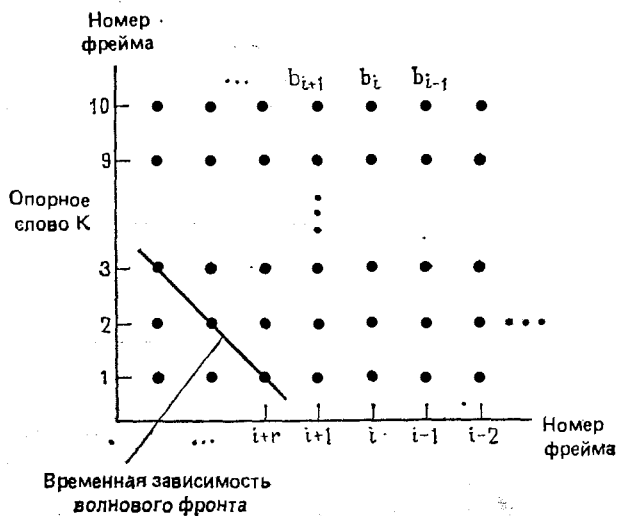
где L — номер коэффициента ЛПИ.

Локальное расстояние $d_{i,r}$ вычисляется одновременно вдоль всего волнового фронта, показанного на рис. 11.12, *a* по мере поступления каждого нового входного фрейма. Мера совпадения между входными и опорными фреймами определяется путем вычислений суммы локальных расстояний во время прогона фрейма. Расширение или каждый входной фрейм может сжиматься или растягиваться (не более чем в 2 раза), с целью учета индивидуальных особенностей скорости речи. На рис. 11.12, *b* показано, что сумма расстояний выбирается из трех возможностей, которые позволяют производить: сжатие (путь 1); ни расширения, ни сжатия (путь 2); расширение (путь 3). Это выражается уравнением

$$s_{i,r} = d_{i,r} + \min (s_{i-2,r-1} + d_{i-1,r}; s_{i-1,r-1}; s_{i-1,r-2} + d_{i,r-1}). \quad (11.12)$$

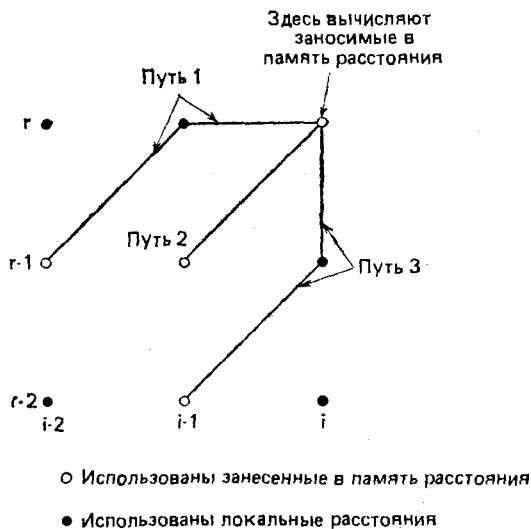
В верхней части столбцов (рис. 11.12, *a*) полученная сумма используется для определения того, закончилось ли слово и было ли выполнено сопоставление с образцом. Идентифицированным считается такое слово в словаре, для которого было получено наилучшее совпадение.

На рис. 11.13 показано, как локальное расстояние и динамическое изменение масштаба времени могли бы быть вычислены с помощью оптического перекрестного процессора. 16 коэффициентов ЛПИ, используемые для обработки очередного входного фрейма, загружаются параллельно в 10 узлов по 16



Непрерывный ввод данных

а



б

Рис. 11.12. Принцип непрерывного динамического изменения масштаба времени.

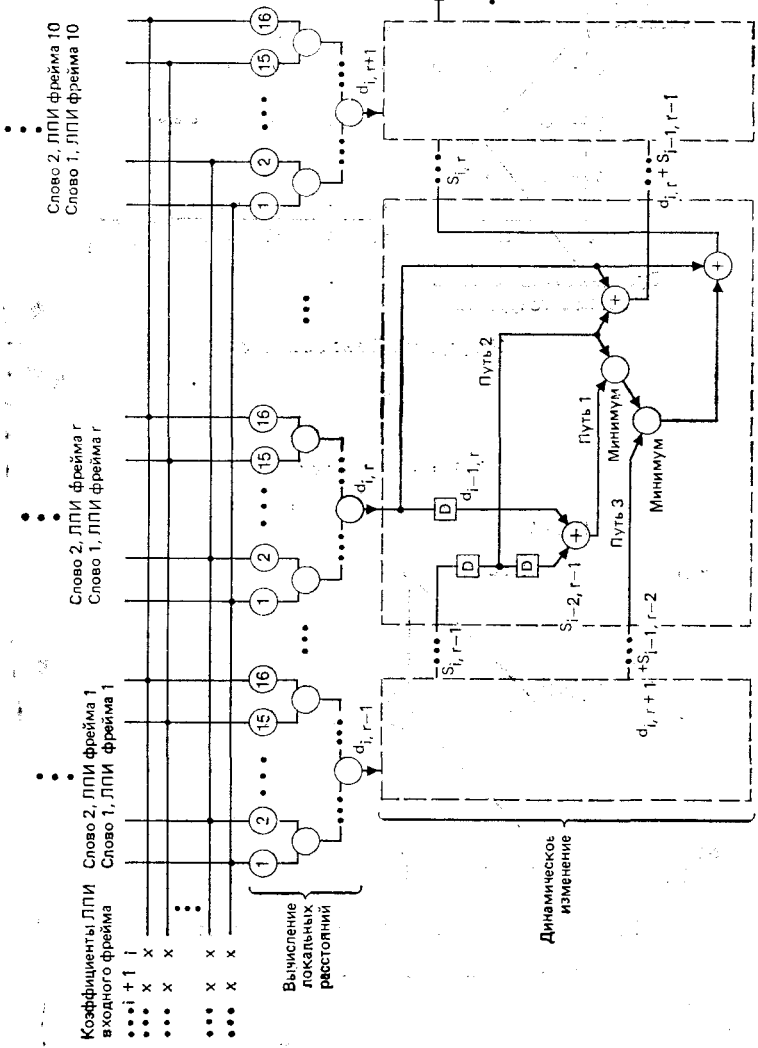


Рис. 11.13. Поточковый граф для алгоритма динамического изменения масштаба времени.

процессоров в каждом и остаются там до тех пор, пока не обработаны все опорные слова. Опорные слова записываются в виде 10 фреймов для коротких слов (0,2 с), 20 фреймов для длинных слов, или 30 фреймов (0,6 с) для очень длинных слов. Это позволяет вводить 10 опорных фреймов посредством 160 входных каналов (как показано на рис. 11.13), а длинные слова — разбивать при этом на две или три части.

В начальном состоянии для простоты рассмотрим одно опорное слово, имеющее 10 фреймов. Все 10 наборов 16-ти коэффициентов ЛПИ для опорного слова вводятся в процессор посредством быстродействующих линий связи с памятью. Древовидные структуры вверху на рис. 11.13 выполняют умножение и суммирование в уравнении (11.11). Опорные фреймы постепенно загружаются таким образом, чтобы обеспечить накопление данных слева направо, по всем фреймам. Выход древовидных структур представляет собой столбец с локальными расстояниями (см. рис. 11.12, *a*), изменяющимися во времени.

Функцию затрат для трех путей, показанных на рис. 11.12, *б*, вычисляют в потоковом графе (рис. 11.13) с помощью уравнения (11.12). Пути обозначены на рис. 11.13. Задержки, обозначенные D , делают возможным использовать фреймы, соответствующие предыдущим моментам времени. Накопленные затраты, показанные с правой стороны, соответствуют величинам, показанным сверху в столбцах на рис. 11.12, *a*, и используются для определения конца слова и идентификации слова.

Корреляция входного фрейма и многочисленных опорных слов выполняется путем конвейерной обработки опорных слов. Все опорные слова обрабатываются в специальной области обработки фреймов системы перед переходом к вычислению корреляции со следующим входным фреймом. Задержки, обозначенные D на рис. 11.13, становятся буферами, в которых результаты для всех опорных слов сохраняются до тех пор, пока не будут востребованы. Поток, выходящий из правой части схемы на рис. 11.13, представляет накопленные затраты на поиск корреляции между входным фреймом и фреймами всех опорных слов.

На рис. 11.14 показано, как окончание слова может быть обнаружено посредством логических выводов. Ввод данных осуществляется из выхода схемы, справа на рис. 11.13. Результаты вычислений разделенных на части слов должны суммироваться, как показано для слова 1, (*a*) и (*б*). Функция затрат должна просмотреть набор логических операций для определения того, закончилось ли слово в данной точке процедуры. Сами правила показаны на рис. 11.14. Прикладываемые усилия должны быть ниже порогового значения, соответствующие определенному интервалу времени суммарные затраты должны быть ниже порога; кроме того, затраты должны быть меньше, чем в соседних

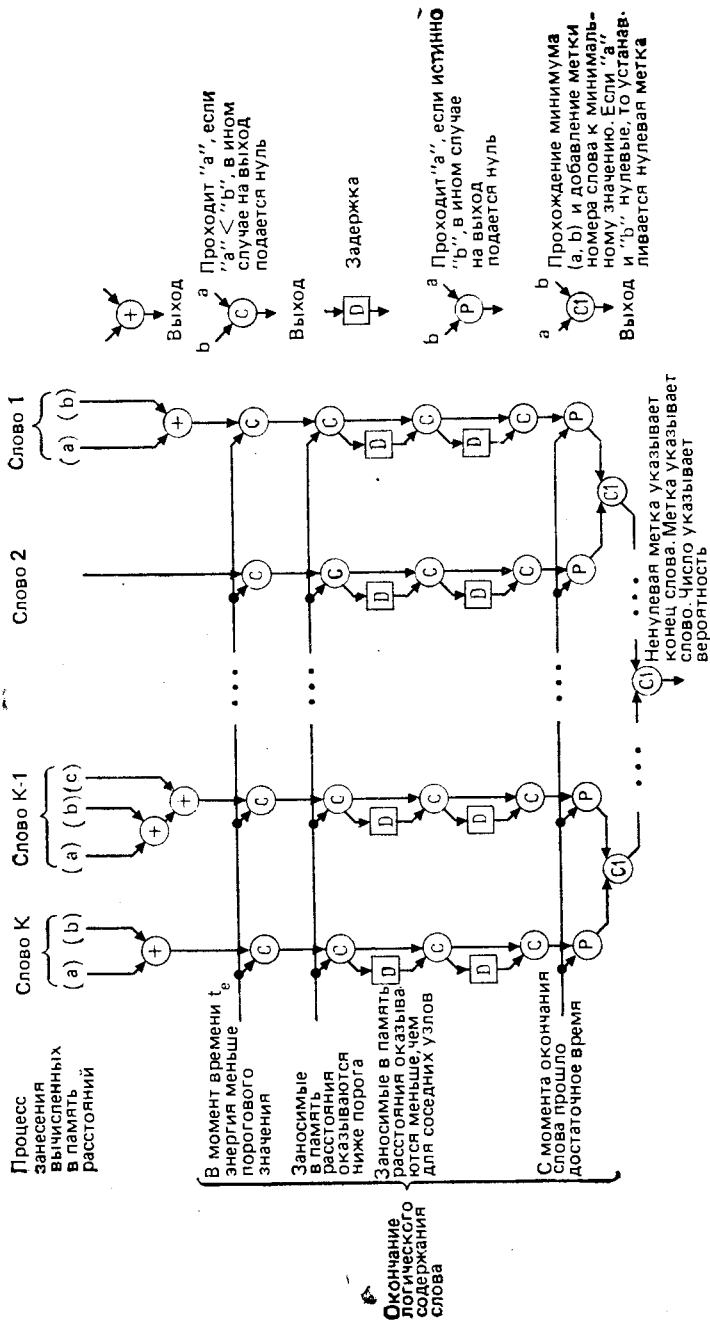


Рис. 11.14. Структурная схема распознавания и обнаружения окончания слова.

точках, и не менее чем определенное число фреймов должно пройти с момента окончания последнего слова. В случае когда слово не соответствует каким-либо тестам, на выход выводится нулевой результат. Далее проводится сравнение между словами. Если слово проходит через минимальный тест, оно будет «отфильтровано» на выход вместе с его значением и идентифицируемым признаком. Наличие ненулевого признака указывает на окончание слова, а само ненулевое значение признака определяет слово. После обнаружения конца слова накопленные функции затрат (в процессе вычисления процедуры динамического изменения временного масштаба) устанавливаются равными нулю, чтобы начать поиск нового слова. Распознанное слово и его возможная принадлежность к какой-либо части речи передаются в анализатор синтаксиса.

11.7.4. Синтаксический анализ речи с помощью экспертной системы синтаксического анализа типа «ситуация — действие», «жди и смотри»

Словарный запас более 10 000 слов будет иметь много сходных слов, и было бы трудно достичь хороших характеристик без использования грамматических правил или информации из контекста, чтобы различить двусмысленные ситуации. Анализатор синтаксиса «жди и смотри» [38, 45] состоит из набора правил и действий, которые требуется выполнить, когда правила запускаются в действие. Приблизительно 500 правил можно было бы реализовать на такой системе, которая показана на рис. 11.2. В качестве иллюстрации рассмотрен набор из 13 правил, описанных в [38].

На рис. 11.15 изображены предложения, подвергаемые синтаксическому анализу, а также получаемое при этом дерево синтаксического анализа. Набор правил, используемый для выработки соответствующего дерева на основе рассматриваемого предложения, описан в [38]. Кроме того, для слов, которые могут относиться более чем к одной части речи, должны рассматриваться все возможные части речи. Если слово относится к более чем одной части речи, то будет необходимо выполнить два пути поиска.

В табл. 11.2 изображено предложение для синтаксического анализа и генерируемое дерево синтаксического разбора. Для этого примера необходимы только два буфера — В1 и В2, в то время как обычно требуются три буфера. Показаны три узла стековой памяти К1, К2 и К3. Буферы и узлы стековой памяти имеют ряд регистров, и обычно содержимое всех регистров перемещается совместно. Слова вводятся слева. Для определения фразы, построенной на основе имени существительного, необходимо использовать предпроцессор. Правила переключаются в

соответствии с содержимым буферов и состоянием верхнего стекового узла K1. Действия, предпринимаемые в результате запуска правила, включают установление связи между буфером В1 и содержимым стека K1, а также сдвиг слов и заполнение буфера В1. Второе действие состоит в создании новой вершины в K1, являющейся началом нового поддерева, то есть глагольной фразы (ГФ). Вершины перемещаются в стеке, чтобы «освободить место» для новой вершины. Последнее действие, используемое в этом примере, предусматривает перемещение вершин из стека K1 в буфер В1 и перемещение вершин в стеке влево для заполнения оставшегося места в стеке K1.

В табл. 11.2 представлен синтаксический разбор предложения, состоящий из 12 шагов [38]. Фраза, содержащая предложение (П), предполагается размещенной в начале в стеке K1, а части речи для анализа предложения (рис. 11.15) вводятся в буферы В1 и В2 слева, начиная с первого слова в предложении. Правило обработки предложения запускается справа тогда, когда K1 содержит вершину предложения и в буфере В1 имеется фраза с существительным (СФ). Предпринимаемые действия состоят в перемещении СФ в K1 и связывании ее с П. В то же время поступающие слова перемещаются вправо для очередного заполнения буфера В1. Другое правило обработки предложения вводится в действие на шаге 2, как показано в табл. 11.2 справа, и это вызывает сдвиг содержимого стека вниз, от K1 до K2, а в стеке K2 размещается новая вершина ГФ. На этой стадии завершено формирование поддерева СФ и начинается построение поддерева ГФ (рис. 11.15). Шаги 4 и 5 связывают ГФ и СФ с поддеревом. Шаг 6 дает начало подде-



Рис. 11.15. Изображение дерева синтаксического анализа.

Таблица 11.2. Примеры шагов анализатора синтаксиса типа «ситуация — действие», «жди и смотри»

Шаг	Буфер				Стек			Правило
	Ожидание	Б2	Б1	К1	К2	К3		
1	СФв	Предл ¹⁾	СФб	Г	СФа	П		Если К1=П, Б1=СФ, П1* тогда подсоединить
2	СФв	Предл	СФб	Г	Г	П(СФа)		Если К1=П(СФ), Б1=Г, тогда образовать ГФ П3
3	СФв	Предл	СФб	Г	Г	ГФ	П(СФа)	Если Х1=ГФ, Б1=Г, тогда подсоединить ГФ2+
4	СФв	Предл	СФб	Предл	СФб	ГФ(Г)	П(СФа)	Если К1=ГФ, Б1=СФ, тогда подсоединить ГФ3
5	СФв	Предл	СФв	Предл	Предл	ГФ(Г, СФб)	П(СФа)	Если Б1=Предл, Б2=СФ и К1=
6	СФв	Предл	СФв	Предл	Предл	ПФ(Г, СФб)	П(СФа)	=ГФ(Г, СФ), тогда образовать ПФ
7	СФв	Предл	СФв	Предл	СФв	ПФ(Г, СФб)	П(СФа)	Если К1=ПФ, Б1=Предл, тогда под- соединить
8	СФв	Предл	СФв	Предл	СФв	ПФ(Г, СФб)	П(СФа)	Если К1=ПФ, Б1=СФ, тогда подсоединить
9	СФв	Предл	СФв	Предл	СФв	ПФ(Г, СФб)	П(СФа)	Если К1=ПФ, Б1=0, тогда К1→Б1
10	СФв	Предл	СФв	Предл	СФв	ПФ(Г, СФб)	П(СФа)	Если К1=ГФ, Б1=ПФ, тогда подсоединить
11	СФв	Предл	СФв	Предл	СФв	ПФ(Г, СФб)	П(СФа)	Если К1=ГФ, Б1=0, тогда К1→Б1
12	СФв	Предл	СФв	Предл	СФв	ПФ(Г, СФб)	П(СФа)	Если К1=П, Б1=ГФ, тогда подсоединить П2

* Образовать фразу предложения + Начинается глагольная фраза † Начинается фраза с предлогом § — Подсоединить ПФ
 К ГФ П — Связать ГФ к П

¹⁾ Предл — сокращение от «предлог».

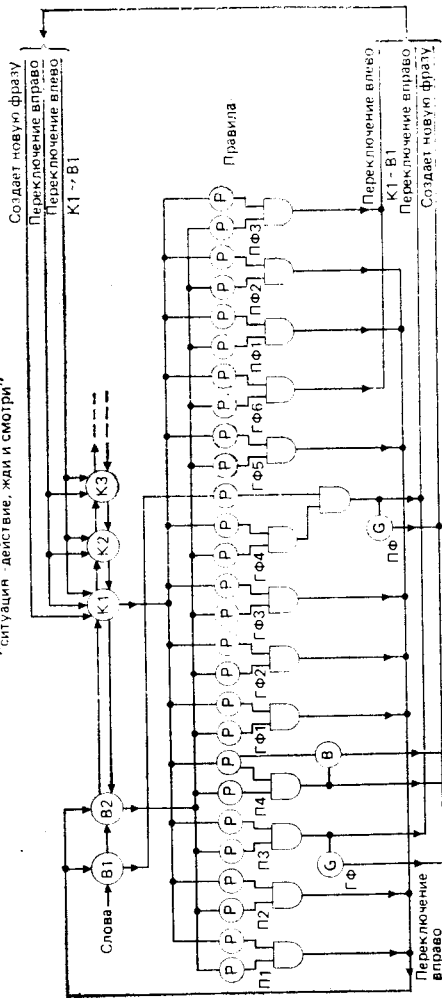
реву фразы с предлогом (ПФ), сдерживая формирование начатых ранее поддеревьев в К3 и К2. Шаги 7 и 8 надстраивают поддерево ПФ. Правило, запускаемое на шаге 9, дает возможность связать ПФ и ГФ, а 10 и 12 шага связывают ГФ с предложением П, что завершает построение полного дерева синтаксического анализа.

На рис. 11.16 для примера, описанного в табл. 11.2 и на рис. 11.15, показан потоковый граф, который может быть реализован в оптическом перекрестном процессоре, аналогично случаю, изображенному на рис. 11.2. Имеется 13 правил с именами, а именно: от S1 до S4, от ГФ1 до ГФ6 и от ПФ1 до ПФ3, имеются входы от буферных регистров и верхнего стекового регистра. За каждый тактовый цикл для каждого из правил сравнивают содержащиеся в нем образцы с подаваемыми на вход сигналами. В данном примере две операции сравнения необходимы для ввода в действие любого правила, кроме как для ГФ4, которое требует трех операций. Также предполагается, что правила не требуют упорядоченного выполнения. Упорядоченные правила могут быть выполнены путем перестроения схемы соединений (как на рис. 11.7). Восемь правил, будучи введенными в действие, вызывают такое переключение выходного триггера, которое вызывает продвижение содержимого буферов, включая перенос данных из В1 вниз, в регистры К1. Правила S3 и ГФ4, если они будут активированы, вызывают следующие операции: сдвиг стековых регистров вправо, генерацию вершины для новой фразы и размещение ее в К1. Правила ГФ6 и ПФ3, будучи запущенными в действие, вызывают сдвиг стековых регистров влево, включая передачу данных из К1 и В1. Запуск правила 4 указывает на завершение синтаксического анализа и вызывает считывание заключительного варианта дерева поиска.

11.8. Выводы

Оптика обладает потенциальной возможностью обеспечить высокий уровень параллелизма обработки и высокую полосу частот, необходимую для реализации сетей межэлементных соединений, требуемых для быстрых, эффективных параллельных вычислений в широком диапазоне символьных и цифровых алгоритмов обработки. ПМС больших размеров и обладающие высокой скоростью могут стать реальностью уже в течение нескольких следующих лет. Была предложена архитектура, использующая ПМС в качестве полностью перестраиваемого перекрестного переключателя. Процессор позволил получить высокую скорость обработки за счет предварительной организации цифровых и символьных потоков данных и обеспечил малые времена прогона. Система показала значительные преимущества по сравнению с альтернативными подходами.

Лотоковый граф для анализа синтаксиса по схеме «ситуация - действие, жди и смотри»

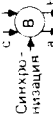


Синхронизация вызывает появление записанного списка на выходе "а".

Логический входной сигнал вызывает изменение в памяти записанного списка в зависимости от значения нового списка с входа "с".

Элемент аналогичен буферу, за тем исключением, что логический сигнал вызывает изменение между собой функций "а" и "с".

Список слов вход



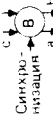
Импульс приходит, если выходной сигнал совпадает с записанным в памяти шаблоном

Генерирует хранящийся в памяти образец, если поступает входной сигнал



Вентиль И

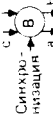
Список слов вход



Импульс приходит, если выходной сигнал совпадает с записанным в памяти шаблоном

Генерирует хранящийся в памяти образец, если поступает входной сигнал

Список слов вход



Импульс приходит, если выходной сигнал совпадает с записанным в памяти шаблоном

Генерирует хранящийся в памяти образец, если поступает входной сигнал

Рис. 11.16. Структурная схема, демонстрирующая синтаксический анализ типа «ситуация — действие», «жди и смотри».

Были подобраны параметры процессора и перекрестных схем соединений, которые обеспечили возможность быстрых и эффективных реализаций символьной экспертной системы, основанной на правилах, цифрового коррелятора, умножителя матрицы на вектор. Реализация алгоритма авторегрессивного моделирования описывалась как иллюстрация к тому факту, что высокая скорость может быть достигнута для широкого диапазона различных алгоритмов.

В заключение был описан в общих чертах символьно-цифровой распознаватель речи, демонстрирующий преимущества одновременно и символьной, и цифровой обработки сигналов. Для всей системы были разработаны потоковые графы, адекватные для отображения на предложенном процессоре. Анализатор синтаксиса использовал правила грамматики для нахождения различий между близкими по смыслу словами, что является трудной задачей для обработки сигналов при наличии очень больших словарных запасов. Необходимы дальнейшие исследования возможностей реализации других анализаторов синтаксиса, которые могли бы различать части речи, соответственно снижая необходимое число слов из словарного запаса, для которых должна повторно выполняться корреляция.

Автор высоко оценивает поддержку данной работы со стороны программы DARPA—ONR, контракт № 00014-85-C-0755. Автор хочет выразить благодарность своим коллегам из Texas Instruments за рассмотрение фрагментов данного текста и высказанные полезные замечания: Э. Боччиери — в вопросах обработки речи; Ч. Хемпхиллу — в вопросах синтаксического анализа, Дж. Сэмпселлу — по вопросам оптики, Д. Оксли — по вопросу организации потоков данных. Автор также хочет поблагодарить Р. Арратуна из Государственного университета им. Уэйна и администрацию Texas Instruments за интерес и поддержку.

Глава 1

1. G. S. Almasi, Magneto optic bubble-domain devices, *IEEE Trans. Magn.*, *MAG 7*: 370 (1971).
2. R. F. Bailey and J. P. Ypma, Bubble domain memories, *Solid State Technology*, *74*, July (1976).
3. J.-P. Krumme, J. Verweel, J. Haberkamp, W. Tolksdorf, G. Bartels, G. P. Espinosa, Thermomagnetic Recording in Thin Garnet Layers, *Appl. Phys. Lett.*, *20*, 11: 451, June (1972).
4. J.-P. Krumme, P. Hansen, K. Witter, Thermomagnetic switching of ferrimagnetic garnet films at their compensation temperature, *J. Appl. Phys.*, *47*, 8: 3681 (1976).
5. J. M. Robertson, S., Wittekoek, Th. J. A. Popma, P. F. Bongers, Preparation and optical properties of bismuth substituted iron garnet films for magneto optic applications, *Appl. Phys.*, *2*: 219 (1973).
6. G. B. Scott, D. E. Lacklison, Magneto optic properties and applications of bismuth substituted iron garnets, *IEEE Trans. Magn.*, *MAG-12*, 4 (1976).
7. D. E. Lacklison, G. B. Scott, R. F. Pearson, J. L. Page, Magneto optic Displays, *IEEE Trans. Magnetics*, *MAG-11*, 5: 1118 (1975).
8. B. Hill, K. P. Schmidt, Thin-Film Iron-Garnet Display Components, SID International Symposium, *Digest of Technical Papers*, Vol. X, Lewis Winner publ., Chicago (1979).
9. B. Hill, K. P. Schmidt, Fast switchable magneto optic memory display components, *Philips J. Res.* *33*: 211 (1979).
10. G. R. Pulliam, W. E. Ross, B. MacNeal, R. F. Bailey, Large stable magnetic domains, *J. Appl. Phys.*, *53* (3): 2754 (1982).
11. W. E. Ross, Demetri Psaltis, H. Anderson, Two-dimensional magneto optic spatial light modulator for signal processing, *Optical Engineering*, *22*, 4: 485 (1983).
12. G. Winkler, Magnetic Garnets, *Vieweg Tracts in Pure and Applied Physics*, *5*, Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden, Germany (1981).
13. P. Hansen, B. Hill, W. Tolksdorf, Optical switching with bismuth-substituted iron garnets, *Philips Techn. Rev.*, *41*, 2: 33 (1983).

14. P. Hansen, Magnetic Anisotropy and Magnetostriction in Garnets, *Physics of Magnetic Garnets*, LXX Corso Soc. Italiana di Fisica, Bologna, Italy (1978).
15. S. Wittekoek, D. E. Lacklison, Investigation of the origin of the anomalous Faraday-rotation of $\text{Bi}_x\text{Ca}_{3-x}\text{Fe}_{3,5+0.5x}\text{V}_{1,5-0.5x}\text{O}_{12}$ by means of the magneto-optic Kerr effect, *Phys. Rev. Lett.*, 28: 740 (1972).
16. B. Hill, Optical transmissivity of Magneto-optic Switching Elements, *Philips J. Res.*, 38: 150 (1983).
17. J.-P. Krumme, V. Doorman, C.-P. Klages, Measurement of the magneto-optic properties of bismuth-substituted iron garnet films using piezobirefringent modulation, *Appl. Optics*, 23, 8: 15, April (1985).
18. V. Doorman, J.-P. Krumme, C.-P. Klages, Measurement of the Refractive Index and Optical Absorption Spectra of Epitaxial Bismuth Substituted Itrium Iron Garnet Films at uv to Near-ir Wavelengths, *Appl. Phys.*, A34, 223-230 (1984).
19. R. Belt, and C. Vemuri, Unpublished work at Litton Airtron (1985).
20. P. Gerard, P. Martin, M. T. Delaye, Some hydrogen implantation effects in bubble garnets, *J. Appl. Phys.* 57: 4058 (1985).
21. B. Hill, H. Meyer, K. P. Schmidt, X-Y Addressed Iron-Garnet Display Components with Integrated Magnetic Control, *Eurodisplay 81*, München Germany, September (1981).
22. B. Hill, X-Y addressing Methods for Iron-Garnet Display Components, *IEEE Trans. Electron Dev.*, ED-27, 9: 1825 (1980).
23. L. J. Gauckler, S. Hofmann, F. Haessner, The Growth of Hillocks and Voids in Gold Single Crystals by Electrotransport, *Acta Metallurgica*, 23: 1541 (1975).
24. M. Etzion, I. A. Blech, Y. Komenm, Study of conductive gold-film lifetime under high current densities, *J. Appl. Phys.*, 46, 4, April (1975).
25. B. Hill, K. P. Schmidt, Light-switching array for high resolution pattern generation, *Philips Techn.*, publication 070, Electronic components and materials: 169 (1982).
26. J. Warner, Nonreciprocal magneto-optic waveguide, *IEEE Trans.*, MTT-23, 1: 70 (1975).

Глава 2

1. H. G. Gibbs, *Optical Bistability, Controlling Light with Light*, Academic Press, New York (1985).
2. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, *Appl. Phys. Lett.*, *45*: 13. (1984). D. A. B. Miller, D. S. Chemla, T. C. Damen, T. H. Wood, C. A. Burrus, A. C. Gossard, and W. Wiegmann, *IEEE J. Quantum Electron.* *OE-21*: 1462 (1985).
3. N. Peyghambarian, to be published in *Opt. Eng.*, invited review article on optical materials for nonlinear devices, February (1986).
4. D. A. B. Miller, D. S. Chemla, D. J. Eilenberger, P. W. Smith, A. C. Gossard, and W. Wiegmann, *Appl. Phys. Lett.*, *42*: 925 (1983).
5. Y. H. Lee, M. Warren, G. R. Olbright, H. M. Gibbs, N. Peyghambarian, T. Venkatesan, J. S. Smith, and A. Yariv, submitted to *Opt. Lett.*
6. Y. Silberberg, P. W. Smith, D. A. B. Miller, B. Tell, A. C. Gossard, and W. Wiegmann, *Appl. Phys. Lett.*, *16*: 701 (1985).
7. D. A. B. Miller, D. S. Chemla, P. W. Smith, A. C. Gossard, and W. T. Tsang, *Appl. Phys. Lett.*, *41*: 679 (1982).
8. H. M. Gibbs, S. S. Targ, J. L. Jewell, D. A. Weinberger, K. Tai, A. C. Gossard, S. L. McCall, A. Passner, and W. Wiegmann, *Appl. Phys. Lett.*, *41*: 221 (1982).
9. S. Ovidia, H. M. Gibbs, J. L. Jewell, D. Sarid, and N. Peyghambarian, *Opt. Eng.*, *24*: 565 (1985).
10. N. Peyghambarian and H. M. Gibbs, *J. Opt. Soc. Am.*, *B 2*: 1215 (1985).
11. M. Dagenais and H. G. Winful, *Appl. Phys. Lett.*, *44*: 574 (1984).
12. N. Peyghambarian, H. M. Gibbs, M. C. Rushford, and D. A. Weinberger, *Phys. Rev. Lett.*, *51*: 1692 (1983).
13. R. Levy, J. Y. Bigot, B. Honerlage, F. Tomasini, and J. B. Grun, *Solid State Commun.*, *48*: 705 (1983).
14. A. K. Kar, J. G. H. Mathew, S. D. Smith, B. Davis, and W. Prettl, *Appl. Phys. Lett.*, *42*: 334 (1983).
15. D. Sarid, R. S. Jameson, and R. K. Hickernell, *Opt. Lett.*, *9*: 159 (1984).
16. C. D. Poole and E. Garmire, *Appl. Phys. Lett.*, *44*: 363 (1984).

17. A. Miller, "Optical Bistability, Dynamical Nonlinearity and Photonic Logic," Proceedings of the Royal Society Meeting on published in Transactions of the Royal Society (1984).
18. F. V. Karpushko and G. V. Sinitsyn, *J. Appl. Spec.*, (USSR) *29*: 1323 (1978).
19. G. Olbright, N. Peyghambarian, H. M. Gibbs, A. Macleod, and F. Van Milligan, *Appl. Phys. Lett.*, *45*: 1031 (1984).
20. S. D. Smith, J. G. H. Mathew, M. R. Taghizadeh, A. Walker, B. S. Wherrett, and A. Hendry, *Opt. Commun.*, *51*: 357 (1984).
21. N. Peyghambarian and H. M. Gibbs, *Opt. Eng.*, *24*: 68 (1985).
22. S. S. Tarng, H. M. Gibbs, J. L. Jewell, N. Peyghambarian, A. C. Gossard, T. Venkatesan, and W. Wiegmann, *Appl. Phys. Lett.*, *44*: 360 (1984).
23. J. L. Jewell, M. C. Rushford, and H. M. Gibbs, *Appl. Phys. Lett.*, *44*: 172 (1984).
24. J. L. Jewell, Y. H. Lee, M. Warren, H. M. Gibbs, N. Peyghambarian, A. C. Gossard, and W. Wiegmann, *Appl. Phys. Lett.*, *46*: 918 (1985).
25. A. Migus, A. Antonetti, D. Hulin, A. Mysyrowicz, H. M. Gibbs, N. Peyghambarian, and J. L. Jewell, *Appl. Phys. Lett.*, *46*: 70 (1985); also see N. Peyghambarian, H. M. Gibbs, J. L. Jewell, A. Migus, A. Antonetti, D. Hulin, and A. Mysyrowicz, Proceedings of the Topical Meeting on Picosecond Electronics and Optoelectronics, Lake Tahoe, Nevada, March (1985).
26. T. Venkatesan, B. Wilkens, M. Warren, Y. H. Lee, G. R. Olbright, H. M. Gibbs, N. Peyghambarian, J. S. Smith, and A. Yariv, *Appl. Phys. Lett.* *48*: 145 (1986).
27. J. L. Jewell and Y. H. Lee, Topical Meeting on Optical Bistability 3, Tucson, Arizona (1985).
28. L. C. West and S. J. Eglash, *Appl. Phys. Lett.*, *46*: 1156 (1985).
29. D. A. B. Miller, A. C. Gossard, and W. Wiegmann, *Opt. Lett.*, *96*: 162 (1984).
30. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, *Phys. Rev. Lett.*, *53*: 2173 (1984).
31. F. A. P. Tooley, S. D. Smith, and C. T. Seaton, *Appl. Phys. Lett.*, *43*: 808 (1983).
32. E. Garmire, J. H. Marburger, S. D. Allen, and H. G. Winful, *Appl. Phys. Lett.*, *34*: 374 (1979).
33. J. L. Jewell (unpublished).
34. A. Lohmann, Proceedings of the Topical Meeting on Optical Computing, Lake Tahoe, Nevada, March (1985).

35. K. H. Brenner and A. Huang, Proceedings of the Topical Meeting on Optical Computing, Lake Tahoe, Nevada, March (1985).
36. D. A. B. Miller, *IEEE J. Quantum Electron.*, *QE-17*: 306 (1981).
37. J. P. van der Ziel and M. Ilegems, *Appl. Opt.*, *14*: 2627 (1975).
38. R. L. Thornton, R. D. Burnham, and W. Streifer, *Appl. Phys. Lett.*, *45*: 1028 (1984).
39. W. A. Little, *Rev. Sci. Instrum.*, *55*: 661 (1984).
40. J. A. Hermann, *Opt. Commun.*, *44*: 62 (1982).
41. D. F. Walls, C. V. Kunasz, P. D. Drummond, and P. Zoller, *Phys. Rev. A*, *24*: 627 (1981).
42. A. Avizienis, *IRE Transactions on Electronic Computers*, *EC-10*: 389 (1961).
43. B. L. Drake, R. P. Bocker, M. E. Lasher, R. H. Patterson, and W. J. Miceli, to be published in *Opt. Eng.*, *25*: 38 (1986).
44. S. D. Smith, J. G. H. Mathew, M. R. Taghizadeh, F. A. P. Tooley, and A. C. Walker, Proceedings of Optical Bistability 3 Meeting, Tucson, Arizona, December (1985).

Глава 3

- 1 Aspnes, D. E. and Bottka, N. (1972). Electric-field on the dielectric function of semiconductors and insulators, *Semiconductors and Semimetals*, Vol. 9 (R. K. Willardson and A. C. Beer, eds.), Academic Press, New York.
- 2 Aull, B. F., Burke, B. E., Nichols, K. B., and Goodhue, W. D. (1987). "Multiple-Quantum-Well CCD Spatial Light Modulators," Proceedings of SPIE: Spatial Light Modulators and Applications II, Vol. 825, San Diego, CA.
- 3 Becker, R. A., Woodward, C. E., Leonberger, F. J., and Williamson, R. C. (1984). Wide-band electrooptic guided-wave analog-to-digital converters, *Proc. IEEE*, *72*: 802.
- 4 Blossy, D. F. and Handler, P. (1972). Electroabsorption, *Semiconductors and Semimetals*, Vol. 9 (R. K. Willardson and A. C. Beer, eds.), Academic Press, New York.
- 5 Bocker, R. P., Caulfield, H. J., and Bromley, K. (1983). Rapid unbiased bipolar incoherent calculator cube, *Applied Optics*, *22*: 804.

- 6Burke, B. E., Goodhue, W. D., and Nichols, K. B. (1986). "CCD spatial light modulators using electroabsorption effects in multiple quantum wells," *Proceedings of SPIE: Nonlinear Optics and Applications*, Vol. 613, Los Angeles, California, pp. 124-129.
- 7Chemla, D. S., Damen, T. C., Miller, D. A. B., Gossard, A. C., and Weigmann, W. (1983). Electroabsorption by Stark effect on room-temperature excitons in GaAs/GaAlAs multiple quantum well structures, *Appl. Phys. Lett.*, 42: 864.
- 8Dereniak, E. L. and Crowe, D. G. (1984). *Optical Radiation Detectors*, Wiley, New York, p. 202.
- 9Deyhimi, I., Eden, R. C., and Harris, J. S. (1980). GaAs and related hetero-junction charge-coupled devices, *IEEE Trans. Electron Devices*, ED-27: 1172.
- 10Fisher, A. D. (1985). "A review of spatial light modulators," *Digest of Topic Meeting on Optical Computing*, Incline Village, Nevada, pp. TuCl-1-TuCl-4.
- 11Goodhue, W. D., Burke, B. E., Nichols, K. B., Metze, G. M., and Johnson, G. D. (1986). Quantum-well charge-coupled devices for charge-coupled device-addressed multiple-quantum-well spatial light modulators, *J. Vac. Sci. and Tech. B*, 4: 769.
- 12Goodhue, W. D., Burke, B. E., Aull, B. F., and Nichols, K. B. (1988). MBE-grown spatial light modulators with charge-coupled device addressing, *J. Vacuum Sci. Tech.* (submitted).
- 13Higgins, J. A. (1984). "Electrooptical signal processing applications of the GaAs CCD," *Proceedings of SPIE: Real Time Signal Processing VII.*, Vol. 495, San Diego, California, pp. 197-204.
- 14Kingston, R. H. (1979). Electroabsorption in GaInAsP, *Appl. Phys. Lett.*, 34: 744.
- 15Kingston, R. H., Burke, B. E., Nichols, K. B., and Leonberger, F. J. (1982). Spatial light modulation using electroabsorption in a GaAs charge-coupled device, *Appl. Phys. Lett.*, 41: 413.
- 16Kingston, R. H. and Leonberger, F. J. (1983). Fourier transformation using an electroabsorptive CCD spatial light modulator, *IEEE J. Quantum Electronics*, QE-19: 1443.
- 17Lindley, J. P. (1982). Applications of acousto-optic techniques to rf spectrum analysis, *Acousto-Optic Signal Processing* (N. J. Berg and J. N. Lee, eds.), Dekker, New York.
- 18Milano, R. A., Liu, Y. Z., Anderson, R. J., Sovero, E. A., and Cohen, M. J. (1981). "Observation of two-dimensional charge transfer in GaAs," *IEDM Tech. Digest*, Washington, DC. pp. 626-628.

- 19Milano, R. A., Liu, Y. Z., Anderson, R. J., and Cohen, M. J. (1982). Very low dark current heterojunction CCDs, *IEEE Trans. Electron Devices*, ED-29: 1294.
- 20Nichols, K. B. and Burke, B. E. (1985). A gallium arsenide overlapping-gate charge-coupled device, *IEEE Electron Device Lett.*, EDL-6: 237.
- 21Nichols, K. B., Burke, B. E., Aull, B. F., Goodhue, W. D., Gramstorff, B. F., Hoyt, C. D., Vera, A. (1988). Spatial light modulators using charge-coupled-device addressing and electroabsorption effects in GaAs/AlGaAs multiple quantum wells, *Appl. Phys. Lett.*, 52: 1116.
- 22Patterson, R. H., Bromley, K., Monahan, M. A., and Symanski, J. J. (1984). "A real-time incoherent optical signal processor system," Proceedings of SPIE, Real Time Signal Processing VII, Vol. 495, San Diego, California, pp. 178-189.
- 23Sahai, R., Pierson, R. L., Jr., Martin, E. H., and Higgins, J. A. (1984). High speed GaAs detector array/CCD multiplexer for acousto-optic spectrum analyzer," Proceedings of SPIE, Optical Technology for Microwave Applications, Vol. 477, Arlington, Virginia, pp. 165-173.
- 24Schroder, D. K. (1984). Schottky barrier gate charge-coupled devices, *Metal-Semiconductor Schottky Barrier Junctions and Their Applications* (B. L. Sharma, ed.), Plenum Press, New York.
- 25Sequin, C. H. and Tompsett, M. F. (1975). *Charge Transfer Devices*, Academic Press, New York.
- 26Sovero, E., Hill, W. A., Sahai, R., Higgins, J. A., Pittman, S., Martin, E. H., and Pierson, R. L., Jr. (1983). "Transversal filter application of a high-speed gallium arsenide CCD," Proceedings GaAs IC Symposium, Phoenix, Arizona, pp. 92-95.
- 27Sovero, E. A., Sahai, R., Hill, W. A., and Higgins, J. A. (1984). "Microwave frequency GaAs charge-coupled devices," Proceedings GaAs IC Symposium, Boston, Massachusetts, pp. 101-104.
- 28Stillman, G. E., Wolfe, C. M., Bozler, C. O., and Rossi, J. A. (1976). Electroabsorption in GaAs and its application to waveguide detectors and modulators, *Appl. Phys. Letters*, 28: 544.
- 29Sze, S. M. (1981). *Physics of Semiconductor Devices*, Wiley, New York, p. 248.
- 30Tanguay, A. R., Jr., (1984). "Materials requirements for optical processing and computing devices," Proceedings of SPIE: Optical Computing, Vol. 456, Los Angeles, California, pp. 130-158.
- 31Wakita, K., Kawamura, Y., Yoshikuni, Y., Asahi, H. (1985). High-temperature excitons and enhanced electroabsorption in InGaAs/InAlAs multiple quantum wells, *Electronics Lett.*, 21: 574.

- 32 Wood, T. H., Burrus, C. A., Miller, D. A. B., Chemla, D. S., Damen, T. C., Gossard, A. C., and Weigmann, W. (1984). High-speed optical modulation with GaAs/GaAlAs quantum wells in a *p-i-n* diode structure, *Appl. Phys. Lett.*, 44: 16.

Глава 4

1. A. Mukhopadhyay, Complete sets of logic primitives, *Recent Developments in Switching Theory* (A. Mukhopadhyay, ed.), Academic Press, New York (1971).
2. M. Davio, J. Deschamps, and A. Thayse, *Discrete and Switching Functions*, McGraw-Hill, New York (1978).
3. A. W. Burks and J. B. Wright, Theory of logic nets, *Proc. IRE*, 41 (1957).
4. E. L. Post, Introduction to a general theory of elementary propositions, *Am. J. Math.*, 43 (1921).
5. B. A. Bernstein, "Modular Representations of Finite Algebras," Proceedings of the 7th International Congress of Mathematicians I, pp. 207-216 (1928).
6. K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall, Englewood Cliffs, New Jersey (1961).
7. N. S. Szabo and R. I. Tanaka, *Residue Arithmetic and Its Applications to Computer Technology*, McGraw-Hill, New York (1967).
8. D. H. Lehmer, A photo-electric number sieve, *Am. Math. Monthly*, 40: 401 (1933).
9. H. J. Caulfield, J. A. Neff, and W. T. Rhodes, Optical computing: the coming revolution in optical signal processing, *Laser Focus*, 19, 100-110 (1983).
10. A. A. Sawchuck and T. C. Strand, Digital optical computing, *Proc. IEEE*, 72: 758-779 (1984).
11. A. Huang, "The Implementation of a Residue Arithmetic Unit via Optical and Other Physical Phenomena," Proceedings of the International Computing Conference, Washington, D. C., IEEE Cat. No. 75 CH0941-5 C; April (1975).
12. S. A. Collins, Jr., J. Ambuel, and E. K. Damon, "Numerical Optical Data Processing," Proceedings of the International Optical Computing Conference, London, England, IEEE Cat. No. 79CH1305-2, p. 194 (1978).

13. M. T. Fatehi, K. C. Wasmundt, and S. A. Collins, Jr., Optical logic gates using liquid crystal light valve: implementation and application example, *Appl. Opt.*, *20*: 2250-2256 (1981).
14. F. A. Horrihan and W. W. Stoner, Residue-based optical processor, *Proc. SPIE*, *185*: 19-27 (1979).
15. C. C. Guest and T. K. Gaylord, Truth-table look-up optical processing utilizing binary and residue arithmetic, *Appl. Opt.*, *19*: 1201-1207 (1980).
16. A. Huang, Y. Tsunoda, J. W. Goodman, and S. Ishihara, Optical computation using residue arithmetic, *Appl. Opt.*, *18*: 149-162 (1979).
17. D. Psaltis and D. Casasent, Optical residue arithmetic: a correlation approach, *Appl. Opt.*, *18*: 163-171 (1979).
18. A. Tai, I. Cindrich, J. R. Fienup, and C. C. Aleksoff, Optical residue arithmetic computer with programmable computation modules, *Appl. Opt.*, *18*: 2812-2823 (1979).
19. K. Pulliam, Cyclic number codes for optical computing, *Proc. SPIE*, *495*: 29 (1984).
20. J. Jackson and D. Casasent, Optical systolic array processor using residue arithmetic, *Appl. Opt.*, *22*: 2817-2821 (1983).
21. G. Eichmann, J. Keybl, and R. Mammone, Number theoretic transform modular residue processors, *Proc. SPIE*, *209*: 66-72 (1979).
22. J. G. Rosenberg, Completeness properties of multi-valued logic algebras, *Computer Science and Multi-Valued Logic* (D. C. Rine, ed.), North-Holland, New York (1984).
23. M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, U.S. Dept. of Commerce, NBS Applied Math. Series 55, U.S. Govt. Printing Office (1964).
24. J. D. Armitage and A. W. Lohmann, Theta modulation in optics, *Appl. Opt.* *4*: 399-403 (1965).
25. H. Bartelt, A. W. Lohmann, and E. E. Sicre, Optical logical processing in parallel with theta modulation, *J. Opt. Soc. Am. A.*, *1*: 944-951 (1984).
26. P. W. Smith, On the physical limits of digital optical switching and logic elements, *BSTJ*, *61*: 1975-1993 (1982).
27. P. W. Smith, guest editor, Special Issue on Optical Bistability, *IEEE J. Quantum Electr.*, *QE-17* (1981).
28. K. Otsuka, Nonlinear antiresonant ring interferometer, *Opt. Letters*, *8*: 471-73 (1983).
29. G. Eichmann, "Phase Conjugate Optical Four-Wave Mixing Devices for Optical Logic, Computation and Interconnect," post-deadline paper, OSA Topical Meeting on Optical Computing, Lake Tahoe, Nevada (1985).

30. B. H. Soffer, D. Boswell, A. M. Lacker, A. A. Sawchuck, T. C. Strand, and A. R. Tanguay, Jr., Optical computing with variable grating mode liquid crystal devices, *Proc. SPIE*, 232: 128-136 (1980).
31. G. Warde, A. M. Weiss, A. D. Fischer, and J. A. Thackara, Optical information processing characteristics of the microchannel spatial light modulator, *Appl. Opt.*, 20: 2066-2074 (1981).
32. K. Tai, H. M. Gibbs, N. Peyghambarian, and A. Mysyrowitz, Mirrorless dispersive optic bistability due to crosstrapping of counterpropagating beams in sodium vapor, *Opt. Lett.*, 10: 220 (1985).
33. J. Slupecki, A criterion of fullness of many-valued systems of propositional logic, *Studia Logica*, 30: 153-157 (1972). (English translation of the Polish original which appeared in *Compte Rendus des Sciences de la Societe des Sciences et des Lettres de Varservie Cl. II* 32 (1939).

Глава 5

1. Optical Computing, Special Issue, *Proc. IEEE*, 72, 7, July (1984).⁶
2. Optical Computing, Special Issue, *Opt. Eng.*, 23, 1, January-February (1985).
3. K. Julian, Defense program pushes microchip frontiers, *High Technology*, 5: 49-57 (1985).
4. H. Szu, "Neural Network Models for Computing," ONR Optical Processing Meeting, San Diego, California, August (1985).
5. R. A. Athale, "Optical Attentive Associative Memory," Opt. Soc. Am. Meeting, Washington, D. C., October (1985).
6. S. C. Gustafson, J. A. Kirk, G. R. Little, R. P. Kenan, and C. M. Verber, Optical implementations of lumped threshold logic, *Proc. SPIE*, 564, 22 (1985).
7. J. W. Goodman, A. R. Dias, and L. M. Woody, Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms, *Opt. Lett.* 2: 1-3 (1978).
8. S. C. Gustafson, Adaptive optical processor algorithms, architectures, and devices, *Proc. SPIE*, 551, 22 (1985).
9. M. L. Dertouzos, *Threshold Logic: A Synthesis Approach*, MIT Press, Cambridge, Massachusetts (1965).

10. P. M. Lewis II and C. L. Coates, *Threshold Logic*, Wiley, New York, (1967).
11. S. Muroga, *Threshold Logic and Its Applications*, Wiley, New York, (1971).
12. R. Arrathoon and M. H. Hassoun, Optical threshold elements for digital computation, *Opt. Lett.*, 9: 143-145, (1984).
13. A. Pal, A four-variable programmable universal logic module using digital summation threshold logic gates, *Proc. IEEE*, 72: 1813-1815, (1984).
14. L. A. M. Bennett, Realization of logic functions by multi-output threshold-logic gates, *IEE Proc.*, 129E: 239-242, (1982).
15. K. G. Murty, *Linear Programming*, Wiley, New York, (1983).
16. S. Muroga, *Logic Design and Switching Theory*, Wiley, New York (1979).
17. M. H. Hassoun and R. Arrathoon, "Pattern Recognition and Logic with Optical Gates," Opt. Soc. Am. Meeting on Optical Computing, Incline Village, Nevada, March (1985).
18. S. C. Gustafson, "Optical processing research," presented at Innovation Science and Technology Consortium on Optical Computing, Washington, D.C., October (1985).
19. D. Psaltis and N. Farhat, Optical information processing based on an associative-memory model of neural nets with thresholding and feedback, *Opt. Lett.*, 10: 78-100, (1985).
20. D. Psaltis, "Theoretical Investigation of the Effective Utilization of Parallelism and Connectivity in Optical Computers," ONR Optical Processing Meeting, San Diego, California, August (1985).
21. T. Kohonon, *Self-Organization and Associative Memory*, Springer, Berlin, (1984).
22. T. K. Gaylord, M. M. Mirsalehi, and C. C. Guest, Optical digital truth table look-up processing, *Opt. Eng.*, 24: 48-58, (1985).
23. F. T. S. Yu, *White-Light Optical Signal Processing*, Wiley, New York (1985).
24. R. Arrathoon, E. R. Schroeder, and L. D. Hutcheson, Integrated electro-optic Bragg processors for digital real-time signal processing, *Opt. Lett.*, 10: 244-246, (1985).
25. R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical Holography*, Academic, New York, (1971).
26. G. P. McCormick, *Nonlinear Programming: Theory, Algorithms, and Applications*, Wiley, New York (1983).

27. C. C. Guest and T. K. Gaylord, Phase stabilization system for holographic optical data processing, *Appl. Opt.*, 24: 2140-2144 (1985).
28. G. J. Dunning, E. Marom, Y. Owechko, and B. H. Soffer, All Optical Associative Memory with Feedback Using Phase Conjugate Mirrors," Opt. Soc. Am. Meeting, Washington, D.C., October (1985).
29. R. O. Winder, Private communication, February (1985).
30. S. C. Gustafson, D. L. Flannery, and R. O. Winder, Analysis of threshold logic for application to optical computing, *Proc. SPIE*, 564, 24, (1985).
31. S. C. Gustafson, Electro-optical intensity modulation of near-surface light for optical signal processing applications, *Proc. SPIE*, 517: 163-166, (1984).
32. N. Peyghambarian and H. M. Gibbs, Optical bistability for optical signal processing and computing, *Opt. Eng.*, 24: 68-73, (1985).
33. C. Verber, Private communication, March (1985).
34. D. S. Chemla, D. A. B. Miller, and P. W. Smith, Nonlinear optical properties of GaAs/GaAlAs multiple quantum well-material: phenomena and applications, *Opt. Eng.*, 24: 551-564 (1985).
35. S. D. Smith, I. Janossy, H. A. MacKenzie, J. G. H. Mathew, J. J. E. Reid, M. R. Taghizadeh, F. A. P. Tooley, and A. C. Walker, Nonlinear optical circuit elements as logic gates for optical computers: the first digital optical circuits, *Opt. Eng.*, 24: 569-574 (1985).
36. S. D. Smith, Lasers, nonlinear optics and optical computers, *Nature*, 316: 319-324 (1985).
37. J. A. Neff, "Future Directions in Optical Processing," NASA Conf. 2302, Hampton, Virginia, August (1983).
38. A. Huang, "Digital Techniques in Optical Computing," Opt. Soc. Am. Meeting on Optical Computing, Incline Village, Nevada, March (1985).
39. W. T. Rhodes, "Optical Computing Research Architectures and Algorithms," ONR Optical Processing Meeting, San Diego, California, August (1985).
40. R. Kostuk, J. W. Goodman, and L. Hesselink, Optical imaging applied to microelectronic chip-to-chip interconnections, *Appl. Opt.*, 24: 2851-2858 (1985).
41. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Nat. Acad. Sci.*, 19: 2554-2558 (1982).
42. Y. S. Abu-mostafa, "Efficiency of Computation in Neural Networks," Opt. Soc. Am. Meeting, Washington, D.C., October (1985).

Глава 6

- 1 Berg, N. J. and Lee, J. N. (1983). *Acousto-Optic Signal Processing: Theory and Implementation*, Dekker, New York.
- 2 Caulfield, H. J., Neff, J. A., and Rhodes, W. T. (1983). Optical computing: the coming revolution in optical signal processing, *Laser Focus/Electro-Optics*, 19: 100-110.
- 3 Verber, C. M. (1985). Intergrated optical architectures for matrix multiplication, *Optical Engineering*, 24(1): 19-25.
- 4 Abraham, E., Seaton, C. T., and Smith, S. D. (1983). The optical computer, *Scientific American*, 2: 85-93.
- 5 Arrathoon, R. and Hassoun, M. H. (1984). Optical threshold logic elements for digital computation, *Optics Letters*, 9: 143-145.
- 6 Gustafson, S. C. (1984). "Electro-optical intensity modulation of near-surface light for optical signal processing applications," *Proc. SPIE*, 517: 163-166, Boston, Massachusetts.
- 7 Chow, C. K. (1961). "On the Characterization of Threshold Functions," *Proceedings IEEE Symposium on Switching Circuit Theory and Logical Design*, 34-38.
- 8 Hanson, W. H. (1963). Ternary threshold logic, *IEEE Trans. Elec. Computers*, EC-12, (6): 191-197.
- 9 Merrill, R. D. (1964). Some properties of ternary threshold logic. *IEEE Trans. Elec. Computers*, EC-13 (10): 632-635.
- 10 Kitahashi, T., Nomura, H., Tezuka Y., and Kasahara, Y. (1969). The characterizing parameters of ternary logical functions and their applications to threshold logical functions, *Trans. I.E.C.E. Japan*, 52-C (10): 503-510.
- 11 Kitahashi, T. (1970). On the complete monotonicity as the necessary and sufficient condition for a ternary logical function to be a threshold function, *Trans. I.E.C.E. Japan*, 53-C (8): 507-513.
- 12 Aibara, T. and Akagi, M. (1970). Generation of ternary threshold functions of up to three variables, *Systems-Computers-Controls*, 1, (5): 4-12.
- 13 Aibara, T. and Akagi, M. (1972). Enumeration of ternary threshold functions of three variables, *IEEE Trans. Comp.*, C-21: 402-407.

- 14Nazarala, J., and Moraga, C. (1974). "Minimal Realization of Ternary Threshold Functions," *Proceedings 1974 International Symposium on Multiple-Valued Logic*, Morgantown, West Virginia, pp. 347-358.
- 15Blyumin, S. L. (1972). Threshold p -ary functions, *Tekhnicheskaya Kibernetika*, 10(1): 101-108.
- 16Nomura, H. (1972). Some properties of monotonicity on many-valued logic and their applications to the analysis of many-valued threshold function. *Trans. I.E.C.E. Japan*, 55-D (4): 245-251.
- 17Nomura, H. (1973). Characteristic vectors of multivalued logic functions and their application to the realization of multivalued threshold functions, *Systems-Computers-Controls*, 4(1): 18-24.
- 18Fujita, Y., Kitahashi, T., and Tanaka, K. (1970). The functional completeness of many-valued threshold functions, *Trans. I.E.C.E. Japan*, 53-C, (5): 341-342.
- 19Moraga, C. and Schulte-Ontrop, R. (1982). "On Some Cardinality Questions in Multiple-Valued Extended Threshold Logic," *Proceedings 12th International Symposium on Multiple-Valued Logic*, Paris, France, pp. 253-259. IEEE-CS-Press, Silver Spring, Maryland.
- 20Dao, T. (1977). Threshold I^2L and its applications to binary symmetric functions and multivalued logic, *IEEE J. Solid State Circuits*, SC-12: 463-472.
- 21Hurst, S. L. (1986). Multiple-valued logic: its status and its realization, *Optical Engineering*, 25: 44-55.
- 22Cerf, V. G. et al. (1971). Formal control-flow properties of a graph model computation. Report ENG-7178, Computer Science Dept., UCLA.
- 23Moraga, C. (1979). "Extensions on Multiple-Valued Threshold Logic," *Proceedings 9th International Symposium on Multiple-Valued Logic*, Bath, England, pp. 232-240. IEEE-CS-Press, Silver Spring, Maryland.
- 24Moraga, C. (1972). "Non-Linear Ternary Threshold Logic," *Conference Record 1972 Symposium on the Theory and Applications of Multiple-valued Logic*, Buffalo, New York.
- 25Moraga, C. (1977a). "Adaptive Separation in Ternary Switching Circuits," *Proceedings IFAC Symposium on Discrete Systems 5*, Dresden, GDR, pp. 118-127.
- 26Moraga, C. (1977b). A monograph on ternary threshold logic, *Computer Science and Multiple-Valued Logic* (D. Rine, ed.), North-Holland, Amsterdam, Part III.
- 27Hassoun, M. H. and Arrathoon, R. (1985). "Pattern Recognition and Logic with Optical Gates," *Digest of Topical Meeting on Optical Computing*, WB-5, Optical Society of America, Lake Tahoe, Nevada.

Глава 7

1. M. Monahan, K. Bromley, and R. P. Bocker, Incoherent optical correlators, *Proc. IEEE*, 65: 121 (1977).
2. J. W. Goodman, A. R. Dias, and L. M. Woody, Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms, *Opt. Lett.*, 2: 1 (1978).
3. H. M. Gibbs, S. L. McCall, and T. N. C. Venkatesan, Optical bistable devices: the basic components of all-optical systems?, *Opt. Eng.*, 19: 463 (1980).
4. J. W. Goodman, F. J. Leonberger, S.-Y. Kung, and R. A. Athale, Optical interconnections for VLSI systems, *Proc. IEEE*, 72: 850 (1984).
5. H. J. Caulfield, W. T. Rhodes, M. J. Foster, and S. Horvitz, Optical implementation of systolic array processing, *Opt. Comm.*, 40: 86 (1981).
6. D. Casasent, Acousto-optic transducer in iterative optical vector-matrix processors, *App. Opt.*, 21: 1859 (1982).
7. H. J. Whitehouse, and J. M. Speiser, Linear signal processing, *Aspects of Signal Processing with Emphasis on Underwater Acoustics* (G. Taconi, ed.), Reidel, Boston, Massachusetts, p. 698 (1976).
8. C. M. Verber, Integrated-optical approaches to numerical optical processing, *Proc. IEEE*, 72: 942 (1984).
9. D. Psaltis, D. Casasent, D. Neft, and M. Carlotto, Accurate numerical computation by optical convolution, *SPIE*, 232: 151 (1980).
10. H. T. Kung, and C. E. Leiserson, Systolic arrays for VLSI, *Introduction to VLSI* (C. A. Mead, and L. A. Conway, eds.), Addison-Wesley, Reading, Massachusetts (1980).
11. S. Cartwright, and S. C. Gustafson, Convolver-based optical systolic processing architectures, *Opt. Eng.*, 24: 59 (1985).
12. S. Cartwright, New optical matrix-vector multiplier, *App. Opt.*, 23: 1683 (1984).
13. R. P. Bocker, S. R. Clayton, and K. Bromley, Electro-optical matrix multiplication using the twos complement arithmetic for improved accuracy, *App. Opt.*, 22: 2019 (1983).
14. P. S. Guilfoyle, Systolic acousto-optic binary convolver, *Opt. Eng.*, 23: 20 (1984).

15. R. P. Bocker, Optical digital RUBIC (rapid unbiased bipolar incoherent calculator) cube processor, *Opt. Eng.*, **23**: 26 (1984).
16. A. P. Goutzoulis, Systolic time-integrating acousto-optic binary processor, *App. Opt.*, **23**: 4095 (1984).
17. D. Casasent, J. Jackson, and C. Neuman, Frequency-multiplexed and pipelined iterative optical systolic array processors, *App. Opt.*, **22**: 115 (1983).
18. R. A. Athale, W. C. Collins, and P. D. Stilwell, High accuracy matrix multiplication with outer product processor, *Appl. Opt.*, **22**: 368 (1983).
19. W. C. Collins, R. A. Athale, and P. D. Stilwell, Improved Accuracy for an optical iterative processor, *SPIE*, **252** (1982).
20. R. A. Athale, and W. C. Collins, Optical matrix-matrix multiplier based on outer product decomposition, *App. Opt.*, **21**: 2089 (1982).
21. D. Psaltis, and R. A. Athale, High accuracy computation with linear analog optical systems: A critical study, *App. Opt.*, **25**: 3071 (1986).
22. D. Casasent, and B. K. Taylor, Banded-matrix high-performance algorithm and architecture, *App. Opt.*, **24**: 1476 (1985).
23. W. T. Rhodes, and P. S. Guilfoyle, Acousto-optic algebraic processing architectures, *Proc. IEEE*, **72**: 820 (1984).
24. R. Arrathoon, E. R. Schroeder, and L. D. Hutcheson, Integrated electro-optic Bragg processors for digital real-time signal processing, *Opt. Lett.*, **10**: 244 (1985).
25. C. M. Verber, R. P. Kenan, and J. R. Busch, Correlator based on an integrated optical spatial light modulator, *App. Opt.*, **20**: 1626 (1981).
26. R. A. Athale, and H. Q. Hoang, High accuracy matrix multiplication with a magneto-optic spatial light modulator, *SPIE*, **431**: 187 (1983).
27. Clark C. Guest, and Robert Tekolste, Designs and devices for optical bidirectional associative memories, *App. Opt.*, **26**: 5055 (1987).

Глава 8

1. D. Schaefer and J. Fischer, Beyond the supercomputer, *IEEE Spectrum*, **19**, 3: 32 (1982).
2. G. C. Fox and S. W. Otto, Algorithms for concurrent processors, *Physics Today*, p. 50, May (1984).
3. Special issue on optical computing, *Proc. IEEE*, **72**, 7 (1984).

4. A. A. Sawchuk and T. C. Strand, Digital optical computing, *Proc. IEEE*, 72: 758 (1984).
5. D. H. Schaefer and J. P. Strong, III, Tse computers, *Proc. IEEE*, 65: 129 (1977).
6. A. Seko and M. Nishikata, Image Emission Platelet laser for optical information processing, *Appl. Opt.*, 16: 1272 (1977).
7. A. Seko, H. Kobayashi, and K. Shimizu, High speed image information processor using a channel plate image intensifier. *Opt. Eng.*, 19: 229 (1980).
8. R. A. Athale and S. H. Lee, Bistability and thresholding by a new photoconductor-twisted nematic liquid crystal device with optical feedback, *Appl. Opt.*, 20: 1424 (1981).
9. M. T. Fatehi, K. C. Wasmundt, and S. A. Collins, Jr., Optical logic gates using liquid crystal light valve: implementation and application example, *Appl. Opt.*, 20: 2250 (1981).
10. P. Chavel, A. A. Sawchuk, T. Strand, A. R. Tanguay, and B. H. Soffer, Optical logic with variable-grating-mode liquid-crystal devices, *Opt. Lett.*, 5: 398 (1980).
11. B. Clymer and S. A. Collins, Jr., Optical computer switching network, *Opt. Eng.*, 24: 74 (1985).
12. B. K. Jenkins, A. A. Sawchuk, T. C. Strand, R. Forchheimer, and B. H. Soffer, Sequential optical logic implementation, *Appl. Opt.*, 23: 3455 (1984).
13. B. K. Jenkins, P. Chavel, R. Forchheimer, A. A. Sawchuk, and T. C. Strand, Architectural implications of a digital optical processor, *Appl. Opt.*, 23: 3465 (1984).
14. J. Tanida and Y. Ichioka, Optical logic array processor using shadowgrams, *J. Opt. Soc. Amer.*, 73: 800 (1983).
15. Y. Ichioka and J. Tanida, Optical parallel logic gates using a shadowcasting system for optical digital computing, *Proc. IEEE*, 72: 787 (1984).
16. M. T. Fatehi, K. C. Wasmundt, and S. A. Collins, Jr., Optical flip-flops and sequential logic circuits using a liquid crystal light valve, *Appl. Opt.*, 23: 2163 (1984).
17. T. Yatagai, Architectures of Optical Cellular Logic Computers, Abstract of OSA Topical Meeting on Optical Computing, Lake Tahoe, Nevada. March (1985).
18. T. Yatagai, Cellular logic architectures for optical computers, *Appl. Opt.*, 25: 1571 (1986).

19. J. Backus, Can programming be liberated from the von Neumann style? A functional style and its algebra of programs, *Commun. ACM*, 21: 613 (1978).
20. S. H. Unger, A computer oriented toward spatial problems *Proc. IRE*, 46: 1744 (1958).
21. K. Preston, Jr., and L. Uhr, (eds.), *Multi-Computers and Image Processing*, Academic Press, New York (1982).
22. A. Rosenfeld, Parallel image processing using cellular arrays, *Computer*, 16: 14 (1983).
23. K. Preston, Jr., M. J. B. Duff, S. Levialdi, P. E. Norgren, and J. Toriwaki, Basis of cellular logic with some applications in medical image processing, *Proc. IEEE*, 67: 826 (1979).
24. A. Reeves, A systematically designed binary array processor, *IEEE Trans.*, C-29: 278 (1980).
25. K. Batcher, Bistable parallel processing systems, *IEEE Trans.*, C-31: 377 (1982).
26. K. K. Maitra, Cascaded switching networks of two-point flexible cell, *IRE Trans.*, EC-11: 136 (1962).
27. J. Sklansky, General synthesis of tributary switching networks, *IEEE Trans.*, EC-12: 464 (1963).
28. R. A. Short, "Two-Rail Cellular Cascades," Proceedings of Fall Joint Computer Conference, p. 355 (1965).
29. R. C. Minnick, Cutpoint cellular logic, *IEEE Trans. EC-13*: 685 (1963).
30. R. C. Minnick, "Cobweb Cellular Arrays," Proceedings of Fall Joint Computer Conference, p. 327 (1965).
31. R. H. Canaday, "Two-Dimensional Iterative Logic," Proceedings of Fall Joint Computer Conference, p. 343 (1965).
32. F. B. Manning, An approach to highly integrated computer-maintained cellular array, *IEEE Trans.*, C- : 536 (1977).
33. A. Rosenfeld and A. Y. Wu, Reconfigurable cellular computer, *Information and Computer*, 50: 64 (1981).
34. J. von Neumann, *Theory of Self-Reproducing Automata*, The University of Illinois Press (1966).
35. M. J. Flynn, Some computer organizations and their effectiveness, *IEEE Trans.*, C-21: 948 (1972).
36. H. Bartelt and A. W. Lohmann, Hybrid logic processing in parallel, *Appl. Opt.*, 22: 2519 (1983).
37. H. Bartelt, A. W. Lohmann, and E. E. Sicre, Optical logic processing in parallel with theta modulation, *J. Opt. Soc. Amer.*, A-1: 944 (1984).

38. T. Yatagai, Optical space-variant logic-gate array based on spatial encoding technique, *Opt. Lett.*, 11: 260 (1986).
39. J. M. S. Prewitt, Object enhancement and extraction, *Picture processing and Psychopictorics*, Academic Press, pp. 75-149 (1970).
40. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, pp. 267-272 (1971).
41. R. Kirsch, Computer determination of the constituent structure of biological images, *Comput. Biomed. Res.*, 4: 315 (1971).
42. D. Casasent and J. Chen, Nonlinear local image processing using coherent optical techniques, *Appl. Opt.*, 22: 808 (1983).
43. C. C. Guest and T. K. Gaylord, Truth-table look-up optical processing utilizing binary and residue arithmetic, *Appl. Opt.*, 19: 1201 (1980).
44. T. Yatagai, M. Inaba, H. Nakano, and M. Suzuki, Automatic flatness tester for very large scale integrated circuit wafers, *Opt. Eng.*, 23: 401 (1984).
45. A. Vander Lugt, Signal detection by complex filtering, *Trans. IEEE, IT-10*: 139 (1964).
46. A. Huang, "Parallel Algorithms for Optical Digital Computers," Proceedings Tenth International Optical Computing Conference, IEEE Catalog 83CH1880-4, 13 (1983).
47. H. Machida, J. Nitta, A. Seko, and H. Kobayashi, High-efficiency fiber grating for producing multiple beams of uniform intensity, *Appl. Opt.*, 23: 330 (1984).
48. M. Oikawa and K. Iga, Distributed-index planar microlens, *Appl. Opt.*, 21: 1052 (1982).
49. H. Ito, N. Komagata, H. Yamada, and H. Inaba, Novel structure of laser diode and light-emitting diode realised by coaxial transverse junction, *Elec. Lett.*, 20: 577 (1984).

Глава 9

1. H. J. Caulfield, Introduction to analog optical processing and computing, *Proc. Soc. Photo-Opt. Inst. Eng.*, 519: v (1984).
2. R. W. Keyes and J. A. Armstrong, Thermal limitations in optical logic, *Appl. Opt.*, 8: 2549 (1969).

3. P. W. Smith, On the physical limits of digital optical switching and logic elements, *Bell Syst. Tech. J.*, 61: 1975 (1982).
4. L. D. Hutcheson, P. R. Haugen, and A. Husain, Gigabit per second optical chip-to-chip interconnects, *Proc. Soc. Photo-Opt. Inst. Eng.*, 587: 116 (1985).
5. T. K. Gaylord and M. M. Mirsalehi, Truth-table look-up processing: number representation, multilevel coding, and logical minimization, *Opt. Eng.*, 25, 1: 22 (1986).
6. R. Arrathoon, Historical perspectives: optical crossbars and optical computing, *Proc. Soc. Photo-Opt. Inst. Eng.*, 752: 2 (1987).
7. R. Arrathoon, Digital optical computing: possibilities and pitfalls, *Proc. Soc. Photo-Opt. Inst. Eng.*, 564: 108 (1985).
8. R. Arrathoon and S. Kozaitis, Shadow-casting for multiple-valued associative logic, *Opt. Eng.*, 25, 1: 29 (1986).
9. A. A. Sawchuk, B. K. Jenkins, and C. S. Raghavendra, Optical crossbar networks, *Computer*, 20: 50 (1987).
10. J. W. Goodman, A. R. Dias, and L. M. Woody, Fully parallel, high-speed incoherent optical method for performing discrete fourier transforms, *Opt. Lett.*, 2: 1 (1978).
11. R. Arrathoon, Reconfigurable fiber optics and generalized crossbar switches, *Proc. Soc. Photo-Opt. Inst. Eng.*, 835: 338 (1987).
12. R. Arrathoon and S. Kozaitis, Design and performance of a 10 MIP optoelectronic central processing unit, *Proc. Soc. Photo-Opt. Inst. Eng.*, 752: 34 (1987).
13. R. Arrathoon and M. H. Hassoun, Design and performance of a dynamically-programmable high-precision threshold gate, *Proc. Soc. Photo-Opt. Inst. Eng.*, 752: 143 (1987).
14. P. S. Guilfoyle and W. J. Wiley, *Proc. Soc. Photo-Opt. Inst. Eng.*, 639: 102 (1986).
15. R. Arrathoon, E. R. Schroeder, and F. Westervelt, *Proc. Soc. Photo-Opt. Inst. Eng.*, 578: 207 (1985).
16. T. Wang, M. Arshad, and R. Arrathoon, Optically controlled fiber optic logic arrays, *Proc. Soc. Photo-Opt. Inst. Eng.*, 977: (1988).
17. R. Arrathoon, Logic based spatial light modulators, *Proc. Soc. Photo-Opt. Inst. Eng.*, 881: 230 (1988).
18. R. Arrathoon and S. Kozaitis, Architectural and performance considerations for a 10^7 instruction/sec optoelectronic central processing unit, *Opt. Lett.*, 12: 956 (1987).
19. Y. Ichioka and J. Tanida, Optical parallel logic gates using a shadow-casting system for optical digital computing, *Proc. IEEE*, 72: 787 (1984).

20. S. P. Kozaitis and R. Arrathoon, Shadow-casting for direct table look-up and multiple-valued logic, *Appl. Opt.*, 24: 3312 (1985).
21. T. K. Gaylord, M. M. Mirsalehi, and C. C. Guest, Optical digital truth table look-up processing, *Opt. Eng.*, 24, 1: 48 (1985).
22. J. W. Goodman, Fan-in and fan-out with optical interconnections, *Optica Acta*, 32: 1489 (1985).
23. P. R. Haugen, S. Rychnovsky, A. Husain, and L. D. Hutcheson, Optical interconnects for high speed computing, *Opt. Eng.*, 25: 1076 (1986).
24. E. J. McCluskey, Minimization of boolean functions, *Bell Syst. Tech. J.*, 35: 1417 (1956).
25. A. H. Scheinman, A Method for simplifying boolean functions, *Bell Syst. Tech. J.*, 41: 1337 (1962).
26. S. J. Hong, R. G. Cain, and D. L. Ostapko, MINI: A heuristic approach for logic minimization, *IBM J. of Res. and Dev.*, 18: 443 (1974).
27. R. Brayton, G. D. Hachtel, L. Hemachandra, A. R. Newton, and A. L. Sangiovanni-Vincentelli, "A Comparison of Logic Minimization Strategies Using ESPRESSO. An APL Program Package for Partitioned Logic Minimization," Proceedings of Int. Symp. on Circ. and Syst. 43, Rome (1982).
28. T. Sasao, Input variable assignment and output phase optimization of PLAs, *IEEE Trans. on Comp.*, C-33, 10: 879 (1984).
29. T. Sasao, On the optimal design of multiple-valued PLAs, *ISMVL-86*: 214 (1986).
30. Z. Kohavi, *Switching and Finite Automata Theory* (McGraw-Hill, New York, p. 209 (1978).
31. R. Arrathoon and M. H. Hassoun, "Optical Threshold Logic Elements for Digital Computation," *Opt. Lett.* 9, 143 (1984).
32. R. Arrathoon, M. Rana, and T. Wang, "On the Limits of Fiber Optic Fan-in and Fan-out for Optoelectronic Logic," *Proc Soc. Photo-Opt. Inst. Eng.*, 936, 327 (1988).

Глава 10

1. C. L. Forgy et al., "Opportunities for Parallelism in AI Tasks," Carnegie-Mellon Reports (1985).
2. H. Winston, *Artificial Intelligence*, Addison-Wesley, Reading, Massachusetts (1984).

3. A. Barr and E. Feigenbaum et al., eds. *The Handbook of Artificial Intelligence*, 3 vols., Kaufman, Los Altos, California (1981).
4. E. Charniak and D. McDermott, *Introduction to Artificial Intelligence*, Addison-Wesley, Reading, Massachusetts (1985).
5. R. D. Fennell and V. R. Lesser, Parallelism in artificial intelligence: a case study of hearsay II, *IEEE Trans. Computers*, C-26: 98-111 (1977).
6. D. Marr, *Vision*, Freeman, San Francisco (1982).
7. R. Lindsay, B. Buchanan, E. Feigenbaum, and J. Lederberg, *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project*, McGraw-Hill, New York (1980).
8. J. McDermott, R1: The formative years, *AI Magazine*, 2: 21 (1981).
9. E. H. Shortliffe, *Computer-Based Medical Consultations: MYCIN*, American Elsevier, New York (1976).
10. F. Hayes-Roth, D. Lenat, and D. Waterman, *Building Expert Systems*, Addison-Wesley, Boston, Massachusetts (1983).
11. H. L. Andrews, Speech processing, *IEEE Computer*, 17: 315 (1984).
12. R. Balzer and L. Erman et al., "HEARSAY-III: a domain independent framework for expert systems," Proc. AAAI-1 (1980).
13. D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, New Jersey (1982).
14. R. D. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York (1973).
15. F. Itakura, Minimum prediction residual principle applied to speech recognition, *IEEE Trans. Acoustics, Speech, and Signal Processing*, ASSP-23: 67 (1975).
16. See, for example, references in *Artificial Intelligence*, 17: (1D3) (Special Issue on Vision) 1981.
17. K. Fu and A. Rosenfeld, Pattern Recognition and computer vision, *IEEE Computer*, 17: 274 (1984).
18. E. Rich, *Artificial Intelligence*, McGraw-Hill, New York (1983).
19. R. Davis and D. Lenat, *Knowledge-Based Systems in Artificial Intelligence*, McGraw-Hill, New York, (1982).
20. See, for example, T. A. Blaxton and B. G. Kushner, An organizational framework for building adaptive artificial intelligence systems, in preparation.
21. B. G. Buchanan, D. H. Smith et al., Applications of artificial intelligence for chemical inference XXII: automatic rule formation in mass spectrometry by means of the MetaDENDRAL program, *J. Am. Chem. Soc.*, 98: 6168 (1976).

22. D. Lenat, EURISKO: a program that learns new heuristics and domain concepts, *Artificial Intelligence*, 21: 31 (1983).
23. J. R. Anderson, *The Architecture of Cognition*, Harvard Univ. Press, Cambridge, Massachusetts (1983).
24. K. Hwang and F. A. Briggs, *Computer Architecture and Parallel Processing*, McGraw-Hill, New York (1984).
25. C. L. Seitz, Concurrent VLSI architectures, *IEEE Transactions on Computers*, C-33: No. 12 (1984).
26. J. A. Neff, "Optical Interconnections Between Integrated Circuit Chips," Proceedings of the International Electronics Packaging Conference, Orlando, Florida (1985).
27. W. D. Hillis, *The Connection Machine*, The MIT Press (1985).
28. D. M. Pepper, Nonlinear optical phase conjunction, *Optical Engineering*, 21: 156 (1982).
29. A. A. Sawchuk, B. K. Jenkins, C. S. Raghavendra, and A. Varma, "Optical Interconnection Networks," Proceeding of the IEEE International Conference on Parallel Processing, St. Charles, Illinois (1985).
30. H. J. Caulfield, W. T. Rhodes, M. J. Foster, and S. Horvitz, Optical implementation of systolic array processing, *Opt. Commun.*, 40: 86 (1981).
31. P. S. Guilfoyle, Systolic acousto-optic binary convolver, *Opt. Eng.*, 23: 20 (1984).
32. S. Cartwright and S. C. Gustafson, Convolver-based optical systolic processing architectures, *Opt. Eng.*, 59 (1985).
33. John Perry, Image understanding by computers: algorithms and implementations, BDM Technical Report BDM/W-86-0131-TR (Feb. 1986).
34. J. L. Baer, Computer architecture, *IEEE Computer*, 17: 77 (1984).

Глава 11

1. A. D. McAulay, "The role of parallel computing", IEEE Region 5 Conf., 85CH2123-8 (1985).
2. A. D. McAulay, Parallel arrays or vector machines, which direction in VLSI?" IEEE International Workshop on Computer Systems Organization, IEEE Comp. Soc., p. 179-184 (1983).
3. A. D. McAulay, Prestack inversion with plane layer point source modeling, *Geophysics*, 50, 77-89 (1985).

4. A. D. McAulay, Plane-layer prestack inversion in the presence of surface reverberation, *Geophysics*, 51, 9 (1986).
5. J. W. Goodman, F. J. Leonberger, S. Y. Kung, and R. A. Athale, Optical interconnections for VLSI systems, *Proc. IEEE*, 72: 850-866 (1984).
6. J. A. Neff, "Electro-optic techniques for VLSI Interconnect," AGARD-NATO Avionics Panel Specialists' Meeting on Digital Optical Circuit Technology (1984).
7. A. D. McAulay, Deformable mirror nearest neighbor optical computer, *Optical Eng.*, 25: 76 (1986).
8. D. Psaltis and N. Farhart, Optical information processing based on an associative-memory model of neural nets with thresholding and feedback, *Optics Letters*, 10, 98-100 (1985).
9. A. D. McAulay, Finite Element Computation on Nearest Neighbor Connected Machines, *NASA Symposium on Advances and Trends in Structures and Dynamics*, Nasa Publ. 2335, p. 15-29 (1984).
10. H. T. Kung, Why systolic architectures? *Computer*, 15: 37-46 (1982).
11. H. T. Kung, High performance microprocessor to be used in groups of hundreds, *IEEE EASCON* (1983).
12. H. T. Kung and O. Menzilcioglu, Warp: A programnable systolic array processor *SPIE*, 495: 130-136 (1985).
13. P. J. Kuekes and M. S. Schlausker, "A One-Third Gigaflop Systolic Linear Algebra Processor," *SPIE*, Vol. 495, Real Time Signal Processing Conf., 137-139 (1984).
14. D. W. Yen et al., ESL Systolic Processor for Signal and Image Processing, *IEEE Comp. Soc. Wksp. Comp. Arch. for Pattern Analysis and Image Database Mangt.* (1981).
15. J. G. Nash and C. Petrozolin, "VLSI Implementation of a Linear Systolic Array," *IEEE Intern. Conf. on Acoustics Speech and Signal processing*, 85CH2118-8 Vol. 3, p 1392-1395 (1984).
16. J. J. Symanski, NOSC systolic processor testbed, *Tech. Rpt., NOSC TD 558* (1983). Progress on a Systolic Processor Implementation, *SPIE 341* (1982).
17. J. M. Speiser and H. J. Whitehouse, "A Review of Signal Processing with Systolic Arrays," *SPIE*, Real Time Signal Processing Conf., Vol. 431, p 2-6 (1983).
18. H. J. Siegel, *Interconnection Networks for Large Scale Parallel Processing, Theory and Case Studies*, Lexington Books (1984).
19. D. Y. Cheng, "A floating point coprocessor for the Butterfly Multi-processor System", SRC Technical report No. 059, Univ. California, Berkeley (1984).

20. J. C. Browne, Parallel architectures for computer systems, *Physics Today*, 37 (1984).
21. C. L. Seitz, The cosmic cube, *Communications of the ACM*, 28: 22-23 (1985).
22. A. A. Sawchuk, B. K. Jenkins, C. S. Raghavendra, and A. Varma, "Optical Interconnection Networks," Proc. IEEE 1985 International Conference on Parallel Processing, 85CH2140-2, p 388-392 (1985).
23. A. D. McAulay, Optical crossbar interconnected signal processor with basic algorithms, *Optical Eng.*, 25: 82 (1986).
24. J. B. Dennis, Dataflow supercomputers, *Computer*, 13: 48-56 (1980).
25. D. R. Treleaven, R. Brownbridge, and R. P. Hopkins, Data-driven and demand driven computer architectures, *Computing Surveys*, 14 (1982).
26. D. Oxley, B. Sauber, and M. Cornish, Software development for dataflow machines, *Handbook of Software Engineering* (C. R. Vick and C. V. Ramamoorthy (eds.) (1984).
27. Arvind and R. A. Iannucci, "Two fundamental issues in multiprocessing: the dataflow solution", MIT Report, MIT/LCS/TM-241 (1983).
28. J. R. Gurd, C. C. Kirkham, and I. Watson. The Manchester prototype data flow computer, *Communication of the ACM*, 28: 34-52 (1985).
29. K. Hiraki, T. Shimada, K. Nishida, "A Hardware Design of the Sigma-1, a Dataflow Computer for Scientific Computations," Proceedings of IEEE International Conf. on Parallel Processing, 84CH2045-2 (1984).
30. M. Cornish et al., "The TI Data Flow Architecture: The Power of Concurrency for Avionics," Proceedings of 3rd Conference on Digital Avionic Systems, p 19-25 (1979).
31. D. R. Pape, Optically addressed membrane spatial light modulators, *Opt. Eng.*, 24: 107-110 (1985).
32. Y. Suematsu, Advances in semiconductor lasers, *Physics Today*, 38: p 32-39 (1985).
33. R. N. Thomas, The mirror-matrix tube: A normal light valve for projection displays, *IEEE Trans.*, ED-22, 9 (1975).
34. D. R. Pape and L. J. Hornbeck, Characteristics of the deformable mirror device for optical information processing, *Opt. Eng.*, 22: 675-681 (1983).
35. J. Allen, Computer architectures for digital signal processing, *IEEE Proc.*, 73: 852-873 (1985).
36. A. D. McAulay, "Conjugate Gradients on Optical Crossbar Interconnected Multiprocessor," Second SIAM Conference on Parallel Processing for Scientific Computation (1985), to appear in *J. Parallel*

- Distributed Comput. (Feb. 1989).
37. W. B. Ackerman, Data flow languages, *Computer*, 15: 15-25 (1982).
 38. P. H. Winston, *Artificial Intelligence*, 2nd ed. Addison-Wesley (1984).
 39. A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*, Prentice-Hall (1975).
 40. A. D. McAulay, "Predictive Deconvolution of Seismic Array Data for Inversion," IEEE International Conference on Acoustics, Speech, and Signal Processing, 85CH2118-8, Vol. 1, p. 180-183 (1985).
 41. S. Y. Kung and Y. H. Hu, A highly concurrent algorithm and pipelined architecture for solving Toeplitz systems, *IEEE Trans Acoustics Speech and Signal Processing*, ASSP-31, p. 66 (1983).
 42. A. D. McAulay, "Parallel AR Computation with a Reconfigurable Signal Processor," IEEE International Conference on Acoustics, Speech, and Signal Processing, 86CH2243-4; Vol. 2 (1986).
 43. A. D. McAulay, "Optimal Least Square Filtering with a Digital Optically Interconnected Processor," SPIE Technical Symposium SE, Advances in Optical Information Processing II, Vol. 639 (1986).
 44. M. P. Marcus, "Image texture classification with an optical crossbar interconnected processor," Army Research Office Workshop on Future Directions in Computer Architecture and Software (1986).
 45. M. P. Marcus, A theory of syntactic recognition for natural language, MIT Press (1980).

ПРИМЕЧАНИЯ К СПИСКУ ЛИТЕРАТУРЫ

Глава 2

1. Х. Гиббс, Оптическая бистабильность. Управление светом с помощью света. Пер. с англ., М.: Мир (1988).
18. Ф. В. Карпушко, Г. В. Синицын, *Журнал прикладной спектроскопии*, т. 29, с. 1323 (1978).

Глава 3

3. Р. А. Беккер, Ч. Э. Вудворт, Ф. Дж. Леонберджер, Р. С. Уильямсон, Широкополосные электрооптические волноводные аналого-цифровые преобразователи, *ТИИЭР*, т. 72, № 7, с. 58 (1984).

Глава 5

9. М. Дертоузос, Пороговая логика, М, Мир (1967).
1. Оптические вычисления, тематический выпуск, *ТИИЭР*, т. 72, № 7 (1984).

Глава 7

4. Дж. У. Гудмен, Ф. Дж. Леонберджер, Г. Суньюань, Р. А. Атхале, Оптические связи в СБИС-системах, *ТИИЭР*, т. 72, № 7, с. 114 (1984).
8. К. М. Вербер, Интегрально-оптические методы численной оптической обработки данных, *ТИИЭР*, т. 72, № 7, с. 218 (1984).
23. У. Т. Родес, П. С. Гилфойл, Архитектура акустооптических алгебраических процессоров, *ТИИЭР*, т. 72, № 7, с. 80 (1984).

Глава 8

3. Оптические вычисления, тематический выпуск, *ТИИЭР*, т. 72, № 7, (1984).
15. Е. Итиока, Дж. Танида, Оптические параллельные логические вентили на основе теневой системы для оптических цифровых вычислителей, *ТИИЭР*, т. 72, № 7, с. 42 (1984).

Глава 9

19. Е. Итиока, Дж. Танида, Оптические параллельные логические вентили на основе теневой системы для оптических цифровых вычислителей, *ТИИЭР*, т. 72, № 7, с. 42 (1984).

Глава 10

5. Дж. Гудмен, Ф. Дж. Леонберджер, Г. Суньюань, Р. А. Атхале, Оптические связи в СВИС-системах, *ТИИЭР*, т. 72, № 7, с. 114 (1984).

- Адресация перекрестная 39
- Архитектура вычислений временной свертки с временным интегрированием 189
- — — — — пространственным интегрированием 189
- со многими потоками команд и одним потоком данных (МКОД) 224
- — — — — и многими потоками данных (МКМД) 224
- с одним потоком команд и многими потоками данных (ОКМД) 224
- — — — — и одним потоком данных (ОКОД) 224
- фон Неймана 218
- чисто оптическая 346
- Архитектуры вычисления свертки оптическими методами 186
- гибридные второго поколения 269
- База знаний 321
- Бистабильные резонаторы *Фабри — Перо* 53
- Время переноса заряда 93
- Выбеливание 105
- Вывод логический 289
- Вычисление разности в кадре 80
- Вычисления символьные 271
- Вычислительные возможности 72
- Гиперповерхность 177
- Граф направленный 284
- Группа *Абелева* р-элементарная 134
- Декодер 240
- Декодер высших порядков 257
- Домены цилиндрические магнитные 14
- Запоминающее устройство динамическое с произвольной выборкой (ДЗУПВ) 253
- — статическое с произвольной выборкой (СЗУПВ) 253
- Знания 275
- Зрение техническое 294
- Изменение временного масштаба динамическое 298
- Информация неточная 290
- Каскад двухканальный 221
- *Мэйтри* 222
- с большим числом входов 221
- Компьютеры клеточные 219
- Корреляция 364
- Коэффициент линейного прогнозирования исходов 364
- объединения по входу волоконно-оптических устройств 245
- — — оптических бистабильных устройств 69
- разветвления по выходу волоконно-оптических устройств 245
- — — электронных логических вентилей 340
- Логика клеточная 218
- — голографическая 231
- пороговая р-значная 164
- троичная 163
- программируемая 291
- Матрица *Вандермонда* 118
- клеточная *Минника* 222
- логическая программируемая 240
- — — динамически 266
- — — оптоэлектронная 240
- мажоритарности 221
- микрокомпьютеров 334
- микролинз 235
- систолическая 78
- Метод максимальной энтропии 389
- теневой энтропии 226

- логической минимизации 256
- — — ESPRESSO 256
- — — *Квина — Маккласки* 256
- — — MINI 256
- — — *Шейманна* 256
- Множество полное 115
- слабое полное 116
- Моделирование авторегрессионное 364
- Морфология 300
- Мультимножество 173
- Обработка данных на уровне вентилей 183
- — — — процессора 183
- — — — регистра 183
- конвейерная 350
- параллельная 333
- списков 272
- Ограничение на плотность упаковки 246
- — — — расцениваемую мощность 246
- Операция корреляции 273
- поиска 273
- сравнения 273
- Оптические логические элементы *Фабри — Перо* 59
- Отношение бинарное 133
- *Псалтиса* 210
- центральное 135
- Отношения порядка 133
- эквивалентности нетривиальные 135
- 4-арные 127
- Память ассоциативная 240
- прямого доступа 240
- Параллелизм оптических вычислений 141
- Переключатель перекрестный волоконно-оптический 263
- Переключение терромагнитное 36
- Перекрестная сеть межэлементных соединений 240
- Перенос заряда 86
- ПЗС структура 76
- — — на барьерах *Шоттки* 83
- — — на GaAs 95
- — — со скрытым каналом 85
- — — с резистивным затворным слоем 90
- Поиск перебором в глубину 284
- — — ширину 284
- от фактов к цели 284
- — цели к фактам 284
- Понимание речи 297
- — непрерывной 298
- естественного языка 313

- Поток данных 224
- Прагматика 301
- Предикат 281
- Преобразователь аналого-цифровой 78
- Произведение внешнее 201
- мощности на ширину полосы пропускания 254
- числа межэлементных соединений на ширину полосы пропускания 248
- Производительность вычислений 241
- Пространственный модулятор света на эффекте электропоглощения в ПЗС 98
- Процессор вычисления свертки систолический акустооптический двоичный 195
- матричный систолический оптический 193
- специальных функций 353
- Разделение обобщенное 177
- Распознавание отдельных слов 298
- речи символично-цифровое 393
- Рассуждения 276
- Свертка в *фурье*-плоскости 187
- Семантика 295
- Сети соединений межэлементных 334
- Символьные подстановки 71
- Синтаксис 295
- Система счисления в остаточных классах 117
- экспертная 295
- — основанная на правилах 363
- нейронная 332
- продукции образец — действие 279
- Сложность вычислений 241
- функциональная 261
- Соединения межэлементные оптические плата — плата 237
- — — чип — чип 237
- Спектроанализатор акустооптический 80
- Структуры данных динамические 290
- — — символьные 271
- квантоворазмерные 57
- Таблицы перекодировочные голографические 127
- — — в клеточной логике 233
- Теорема *Ферма* 122
- Ток темновой 89
- Триггер J — K 159
- Три-стабильность 74

- матрицы на вектор 193
- — — матрицу 195
- -сумматор 155
- Ускоритель 352
- Устройство бистабильное оптическое 69
- вычислительное быстродействующее биполярное некогерентное на светоделительном кубе (RUBIC) 198
- запоминающее постоянное (ПЗУ) 240
- — с произвольной выборкой (ЗУПВ) 240
- логическое программируемое пользователем в зависимости от условий эксплуатации (ЛУПУЭ) 253
- на основе деформируемых зеркал 378
- Фактор заполнения 91
- Ферримагнетизм 19

- Фильтрация, согласованная в клеточной логике 235
- Фонетика 300
- Функция логическая элементарная линейная 115
- — — монотонная 115
- — — самодвойственная 115
- — — сохраняющая 0 и 1 115
- разделяющаяся линейная 173
- — мультилинейная 173
- Частота появления ошибок 245
- Число логических выводов в секунду (липе) 272
- Эвристика 283
- Экситоны 57
- Электромиграция 46
- Элементы знания 271
- Эффект *Фарадея* 26
- *Франца—Келдыша* 98
- электропоглощения в квантовых ямах 105

Предисловие редактора перевода	5
Предисловие редактора серии	7
Предисловие	8
Об авторах	12

ЧАСТЬ I. ПРОСТРАНСТВЕННЫЕ МОДУЛЯТОРЫ СВЕТА 14

Глава 1. Магнитооптические модуляторы света. *Б. Хилл* 14

1.1. Введение	14
1.2. Принцип действия магнитооптических модуляторов	15
1.3. Магнитные свойства ферромагнитных гранатов и механизмы переключения	18
1.4. Оптические свойства магнитооптических гранатовых пленок	26
1.5. Конструкция магнитооптических модуляторов	35
1.6. Обсуждение результатов и перспектив	48

Глава 2. Оптические бистабильные устройства и логические элементы. *Н. Пейгамбарян, Дж. Джуэлл* 52

2.1. Введение	52
2.2. Принципы работы бистабильных устройств	53
2.3. Требования к характеристикам нелинейных устройств и материалов	56
2.4. Причины возникновения нелинейных оптических эффектов в полупроводниках	57
2.5. Современное состояние работ по оптическим бистабильным устройствам	58
2.6. Оптические логические элементы <i>Фабри—Перо (ОЛЭФП)</i>	59
2.7. Устройства на собственном электрооптическом эффекте (СЭОУ)	67
2.8. Факторы, ограничивающие предельные величины коэффициентов усиления, разветвления по выходу, объединения по входу	69
2.9. Концепции архитектуры	71
2.10. Предельные параметры и вычислительные возможности ОБУ	72
2.11. Три-стабильность и многозначная логика	74
2.12. Выводы	74

Глава 3. Компоненты оптического процессора на основе приборов с зарядовой связью (ПЗС). *Б. Е. Бёрк, Дж. Хиггинс* 76

3.1. Введение	76
3.2. Примеры электрооптической обработки сигналов (ЭООС)	78
3.3. Принципы работы ПЗС	83
3.4. Применение ПЗС в качестве выходных устройств ЭООС	91
3.5. Адресуемые ПЗС-пространственные модуляторы света (ПЗС-ПМС)	97

3.6. Выводы	111
ЧАСТЬ II. МНОГОЗНАЧНАЯ И ПОРОГОВАЯ ЛОГИКА	113
Глава 4. Многозначная логика в оптических вычислениях. М. Коннер, Д. Эйхманн	113
4.1. Введение	113
4.2. Проблема полноты в двоичной логике	114
4.3. Многозначные логические схемы, основанные на системе счисления в остаточных классах (ССОК)	117
4.4. Оптические многозначные логические ССОК-процессоры	123
4.5. Проблема полноты в многозначной логике	132
4.6. Краткое изложение и выводы	139
Глава 5. Пороговое кодирование и взвешивание в оптических вычислениях. С. С. Густафсон	140
5.1. Введение	140
5.2. Устройства с внешним пороговым кодированием	146
5.3. Устройства с внутренним пороговым кодированием	154
5.4. Заключение	160
Глава 6. Многозначная пороговая логика. К. Морага	162
6.1. Введение	162
6.2. Многозначная пороговая логика	163
6.3. Мультилинейные разделяющиеся функции	172
6.4. Перспективные разработки	177
ЧАСТЬ III. СИСТОЛИЧЕСКИЕ ПРОЦЕССОРЫ И ОПТИЧЕСКИЕ ЛОГИЧЕСКИЕ МАТРИЦЫ	182
Глава 7. Оптическое умножение матриц. С. Кэртрайт	182
7.1. Введение	182
7.2. Цифровая оптика	183
7.3. Рабочие характеристики	204
7.4. Заключение	214
Глава 8. Архитектуры клеточной логики. Т. Ятагаи	216
8.1. Введение	216
8.2. Клеточная логика	218
8.3. Оптическая двоичная логика и архитектура МКМД	225
8.4. Оптическая реализация	229
8.5. Заключение	236
Глава 9. Волоконно-оптические программируемые логические матрицы. Р. Арратун	237
9.1. Введение	237
9.2. Характеристики волоконно-оптических ПЛМ	245
9.3. Проблемы создания логических схем для волоконно-оптических систем	255
9.4. Перестраиваемые волоконные системы	263
9.5. Выводы	267
ЧАСТЬ IV. СИМВОЛЬНЫЕ ВЫЧИСЛЕНИЯ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ	271
Глава 10. Оптика и символьные вычисления. Дж. А. Нефф, Б. Дж. Кашнер	271
10.1. Введение	271
10.2. Что такое искусственный интеллект?	274
10.3. Функциональные возможности символьных вычислений	293
10.4. Архитектуры символьных вычислений	332

Глоссарий	357
Глава 11. Оптические межэлементные соединения для символьной и цифровой обработки в реальном времени. <i>А. Макаулай</i>	364
11.1. Введение	364
11.2. Роль параллелизма и символьной обработки в вычислениях, проводимых в реальном времени	365
11.3. Преимущества перекрестных оптических схем	369
11.4. Оптический процессор с перекрестной схемой	374
11.5. Получение логических выводов с помощью оптического процессора с перекрестной схемой	382
11.6. Цифровая обработка сигналов с помощью оптического мультипроцессора с перекрестной схемой	386
11.7. Случай одновременного выполнения символьных и цифровых вычислений, пример речевой системы	391
11.8. Выводы	404
Литература	407

OPTICAL COMPUTING DIGITAL AND SYMBOLIC

EDITED BY
RAYMOND ARRATHOON

Department of Electrical and
Computer Engineering
Wayne State University
Detroit, Michigan

MARCEL DEKKER, INC.
New York and Basel

ОПТИЧЕСКИЕ ВЫЧИСЛЕНИЯ

Под редакцией Р. Арратуна

Перевод с английского
канд. физ.-мат. наук А. Ю. Быковского

под редакцией
д-ра физ.-мат. наук Ю. А. Быковского



Москва «Мир» 1993